SUPPORTING PILOT PROCEDURE FOLLOWING IN NOMINAL AND OFF-NOMINAL SITUATIONS THROUGH THE USE OF DISPLAYS OF PROCEDURE CONTEXT

A Thesis
Presented to
The Academic Faculty
by
Steven J. Landry

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2004

SUPPORTING PILOT PROCEDURE FOLLOWING IN NOMINAL AND OFF-NOMINAL SITUATIONS THROUGH THE USE OF DISPLAYS OF PROCEDURE CONTEXT

Approved by:

Dr. Julie A. Jacko, Advisor

Dr. François Sainfort

Dr. Brani Vidakovic

May 14, 2004

Acknowledgements

I want to thank everyone who in one way or another participated in the research described herein. Foremost is my deep appreciation to my advisor Dr. Julie Jacko, first for taking me on as a student under somewhat unusual circumstances, and second for her guidance, encouragement, and support. I'd also like to acknowledge the other members of my committee for lending me some of their experience, time, and effort, and for helping me focus my work.

Thanks also goes to the people who participated in the two experiments discussed in this dissertation, and those who helped me secure those participants. The ability of Bill Jones to obtain the volunteer, unpaid participation of airline pilots was nothing short of amazing. Of course, my great appreciation goes to those pilots who gave up some of their valuable time to help out a student. Josh Smith, the GM of the West Valley Flying Club, played an equally big role by allowing me to access both his facility and his club members for the dissertation research. Some of the greatest help comes unexpectedly, which was the case with Patti Andrews, whose initiative resulted in the majority of my participants. The members of the club who volunteered, and those who just made me feel welcome, are also the recipients of my enduring thanks. Research such as this relies on willing volunteers, but getting such enthusiastic participants made my job much easier.

I'd like to thank all those who helped in somewhat less direct ways. All my lab mates over the years have provided invaluable assistance in ways large and small. Whether for commiseration, collaboration, or outright help, you have my sincere appreciation. The contribution of my coworkers at NASA also needs to be acknowledged. Your support, understanding, and encouragement meant more to me than you probably know.

Finally, I’d like to pass along personal thanks to my family for their love and support, and most importantly to Michael Mashtare, Jr., who has put up with more than anyone (including me). It’s often the one closest who bears the brunt of the difficult parts of such an endeavor, and for that you have my humble gratitude and love.

Table of Contents

List of Figures

List of Tables

Summary

This dissertation provides evidence that information displays to support procedure following can aid performance and increase situational awareness and safety. The intent of such displays is to assist operators in not only following operational procedures, but also in comprehending the context of the procedures, enabling them to understand why, when, and how to deviate from the procedures if necessary.

The results of the dissertation research support several important and novel conclusions. First, the results show that the addition of procedure context increases situation awareness and reduces procedure-following errors, which has been shown to be a significant causative factor for accidents in aviation and other domains. Secondly, the results demonstrate that dynamic displays can be used for this purpose. Pilots were able to understand and utilize the information on the displays without additional workload. .

Despite their interest in, and ability to, detect noncompliance, a pilot's ability to comprehend noncompliance appears to be limited. Pilots do not appear be able to interpret the consequences of that noncompliance, suggesting that the design of displays and procedures should assist them in doing so.

In addition, the results demonstrate that pilots attempted to use procedure information even when clearly outside the scope of the procedure. This means that procedures and procedure-support aids should consider operation outside of its normal bounds in their design, rather than only for nominal operation as is currently the case.

Chapter 1

Introduction, Motivation, and Contribution

In this dissertation I intend to show that information displays to support the execution of operational procedures can aid performance and increase situational awareness and safety. I will also provide guidance for the design of such displays. The intent of such displays is to assist operators in not only following operational procedures, but also in comprehending the context of the procedures, enabling them to understand why, when, and how to deviate from the procedures if necessary. An existing design methodology - procedure context – will provide guidance on the content of displays for this purpose (Ockerman, 2000). Although the experimental efforts are focused on aviation procedures, the results have implications for other safety critical socio-technical systems (such as healthcare, manufacturing, nuclear power, and spacecraft operations).

The importance of this problem is reflected in the observation that the control of most systems in safety critical environments is highly proceduralized. There has been a recognition in some fields (including aviation) of the extent to which procedures impact the work domain. Engineers who design flight deck instrumentation consider procedures when developing new avionics. However, such concern is uncommon, and there has been little human factors work on the design and utilization of the procedures themselves until recently (Degani & Weiner, 1994).

Procedures, despite their importance and ubiquity, are typically developed informally, pieced together from diverse requirements and constraints on the system. The result is that the design philosophy behind a procedure, if there is one, is variable from instance to instance. Procedures are also frequently under-defined and limited in scope.

Information about the assumptions made when the procedure was designed is typically lacking. Operators must translate the requirements of the procedure into action and can encounter situations that were not considered by those designing the procedure. There is often no training on when procedures do not apply (and what to do in those situations), and almost no display support is provided to the operator for following or interpreting procedures, or on how to proceed once a procedure has failed.

Yet operators are expected to follow procedure and are faulted if they deviate from them, without considering whether the underlying procedure contributed to the operator's error, or even whether the procedure should have been followed at all.

Motivation

The research that resulted in this work began as an investigation into how to get two aircraft-pilot systems to remain safely close to one another on approach to an airport (Landry & Pritchett, 2000). The dynamics of the aircraft-pilot systems, individually and together, were not very difficult to determine. Yet uncertainty about future position, caused by system failures, errors by pilots, or reactions by pilots to unforeseen events, went beyond the capability of automation. Pilots, however, guided by good procedure, could safely operate the system in these circumstances. The operation of two aircraft close to one another therefore needed to be governed more by procedure than by dynamics or control.

Researchers had developed procedures for this situation (RTCA SC-186, 2002). The proposed solutions, however, were generally dependent upon pilot execution of a very specific procedure in a very specific manner (e.g. Teo and Tomlin, 2003). After a piloted experiment to test procedures for the task (described in Chapter 3) was accomplished, it became apparent that pilots were unlikely to blindly follow procedure, nor did they have an expectation that the other pilot in the system would necessarily adhere to procedure. This lack of reliability has significant consequences for how the system will need to be designed.

In light of these experiences, a review of the literature regarding procedures was undertaken. This review demonstrated that procedures were under-researched and poorly understood in relation to their importance in human-machine systems. Procedures are a large contributor to accidents in a number of domains, including aviation (National Transportation Safety Board, 1994b), the nuclear industry (Trager, 1988), and maritime (Perrow, 1984). Due to such statistics in aviation, there have been calls for more research into procedures (National Transportation Safety Board, 1969).

One dramatic example of the importance of procedures is the Three Mile Island nuclear facility incident in 1979. The conclusion of the report of the President's Commission on the Accident at Three Mile Island (1979) was that the accident was due to human error caused by deficiencies in training for operating the plant under accident conditions, confusing and misleading operating procedures, poor control room displays for operating the plant during emergency situations, and a lack of training on previous accidents.

In another example, pilots of a Gulf Air Airbus 320 crashed following a missed approach to Bahrain International Airport due to non-adherence to operating procedure (Aviation Safety Network, 2003). The pilots performed an unauthorized maneuver to try and shorten their subsequent approach, lost spatial orientation, and crashed. Despite departing from authorized procedures while doing the orbit at too low an altitude, the aircraft still would not have crashed had the operators adhered to the procedure for reacting to the "ground proximity warning system" (GPWS), a system that detects unsafe approach to terrain. The system's alarm went off 11 seconds prior to impact. The pilots responded to the alarm within 3 seconds by stating "gear's up, flaps up" – an indication that they thought the alarm went off due to an unsafe configuration rather than approaching terrain. The procedure for responding to the GPWS is an immediate climb and does not include checking the configuration.

As a final example, the National Transportation Safety Board (NTSB) indicated that insufficient guidance concerning the recognition of when an approach was unstabilized and a lack of indication that part of the pre-landing procedure was not followed contributed to a fatal accident in Little Rock, AR (National Transportation Safety Board, 2001). Numerous other examples exist in the NTSB database.

Despite the prevalence of procedure-related accidents, only a handful of researchers have focused on procedures. Much of the research is focused on reasons for noncompliance to, and ways to increase compliance with, procedures. Most of the remaining research has generally focused on how to present procedures to the operator. The procedure itself is generally taken as correct, with any deviation from the procedure considered an error. When deviations from procedure exist where an accident has occurred, the approach has been to assume that a better presentation of the procedure is warranted.

This is certainly true in some cases. However, noncompliance to procedure is not unusual, and it does not always cause accidents. In fact, noncompliance to procedures is a necessary adaptation to uncertainty in the system – designers of procedures cannot foresee all the situations in which the procedures will operate, and in many situations operators must deviate from procedures to maintain safety. The extent to which this occurs is difficult to determine, since reports are not generally written except in the case of an accident.

Procedure designers almost exclusively design for nominal situations, where the operator does not have to deal with unusual system configurations or task conditions. When such unusual configurations or conditions exist (off-nominal situations), some procedures may not be applicable and may in fact cause accidents if followed. Depending on the criticality of the system, the procedure designer, and the designer of any automated aid for procedure following, would be well-served to create a procedure (or aid) that will provide useful support outside the nominal range of operation of the procedure, or at least not be brittle in the way it fails under those circumstances.

The above concerns reflect the fact that very little theory on procedures has been developed – what they are, how they are used, how they can be designed and supported. Procedures are similar to automation or displays in that they are an aid to accomplishing a task. As such much of the research work that has gone into displays and automation may apply to procedures. Procedures, however, have their own unique features that need to be explicitly considered.

Contribution

This dissertation will contain the beginnings of a theory of procedures – what they are, how they fit into the human-machine system, and how they can be designed and supported. There is currently very little generalizable theory about procedures in the literature. For the most part procedures are taken as is, with shortcomings of procedures and aids addressed for specific applications and for specific problems.

The theory will view procedures as coupling between human and system which arises in the presence of a goal, and which may indicate a safe or efficient method of arriving at the goal. Procedures also provide standardization and predictability in the execution of a task. Traditionally procedures have been viewed simply as the definition of specific interactions with the systems; shortcomings were addressed by attempting to better define these interactions. As such, the interaction of the operator with the procedure, and the procedure with the system, are both similar to the interaction of the operator with an interface or with automation. The research advances made in understanding human-interface and human-automation interactions can then be applied to procedures. An example of this is the use of cognitive walkthrough for evaluating procedures, which mirrors the same technique used for interface evaluation (Novick, 1999).

However, this dissertation will take a larger view of procedures, where procedures also structure the input-output mapping of the system. In this context, the role of the task environment takes on more significance. This environment, considered in the creation of the procedure, is often not provided to the operator implementing the procedure. This is complicated by the ad hoc manner in which some procedures are developed. Some situations in which the operator could find him- or herself have been considered by the procedure designer, but some have not. Without this knowledge, the operator is in a poor position to determine whether the procedure is still valid or not.

Ockerman has described this as “procedure context”, which will be discussed in detail in Chapter 3. Procedure context provides a framework for conveying this missing information to users. Ockerman applied procedure context to the design of aviation checklists; this dissertation will extend procedure context beyond that application.

As will be discussed in Chapter 3, many of the research efforts on procedure have dealt with static procedures and checklists, and in situations in which compliance is desired. This dissertation is novel in that it examines how procedure information is used in operational procedures (a dynamic task), and in situations in which intentional noncompliance is desired. This is the first research effort that studies whether noncompliance to procedure may be a necessary adaptation to a system that has exceeded the conditions envisioned by the procedure designer, and whether knowledge of the context of the procedure can assist the operator in such situations.

An additional under-appreciated aspect of procedures is that they not only direct action, but can also provide information. In aviation, procedures provide standardization across the numerous vehicles in the system. They provide predictability about the future state of the system – predictability that can be critically important to air traffic controllers and pilots. For example, position reporting is a common procedural requirement whereby a pilot must announce over the radio his or her position at designated times. In doing so, the pilot not only provides an update on the position of her or his airplane, but the future position of that airplane can be determined from knowledge of the procedure that pilot is following. Controllers and other aircraft can then safely determine what actions they can take. This dissertation is one of the first efforts to acknowledge and attempt to understand that aspect of procedures.

As mentioned above, this dissertation has implications beyond aviation. As discussed previously (and revisited in Chapter 3), most safety critical cooperative systems utilize procedures and can benefit from a better understanding of how operators in these systems comprehend the context of their procedures.

Organization

Chapter 2 discusses procedures in detail, including a discussion of aviation procedures. Although procedures, both formal and informal, govern nearly all of our interaction with automated and assistive systems, there has been little done to clarify the nature of procedures. Checklists and instructions, as the visible representation of procedures, have garnered the majority of what attention has been paid to the subject. Only through a thorough understanding of the procedures themselves, and the relation between the operator, the procedure, and the system, can a good understanding of how to support procedure following be developed.

Chapter 3 surveys previous, supporting work. Included is a discussion of procedure context, which in this dissertation is applied to the problem of supporting procedure following. A discussion of a body of research on aviation checklists and procedures is also treated in this chapter. Since the experiments described in later chapters include electronic displays, a description of relevant work on such displays is also included. Supporting work regarding the measures used in this dissertation is covered as well. Finally, a previous experiment which provides support for this dissertation is also discussed.

Chapter 4 will detail the theory behind the design of the displays used in the simulator flight experiment described in Chapter 5. The concept of procedure context is applied to instrument approach procedures to generate information displays, which are tested in a piloted simulation experiment.

The results of the flight experiment are discussed in Chapter 6, where the use of enhanced displays which incorporate procedure context elements will be shown to improve situation awareness and safety without increasing workload. A discussion of those results in provided in Chapter 7, with conclusions and recommendations for future work in Chapter 8.

Chapter 2

Procedures

In considering procedures, one often thinks of checklists (or instructions). However, checklists are not the actual procedure; rather, they are an aid designed to provide support to the person executing the procedure. The procedure is the set of steps, acts, or even sub-procedures that one intends to accomplish in order to achieve a goal.

In the context of the aviation tasks discussed in detail in this dissertation, procedures are used to manage the operation of an aircraft within the national airspace system. In general, however, procedures are used to manage the interaction of a system or systems (e.g. a human operator, a vehicle, a control system) with another system or systems (e.g. automation, roadways, a plant) in situations in which the interacting systems are not deterministically connected. In such situations, the range of interactions between the systems is delimited by procedures. Typical reasons for controlling the interaction between systems are to ensure safety, efficiency, standardization, or predictability.

Procedures as Coupling between System and Operator

Because a procedure can be viewed as a filter through which an operator and a system interact, it is instructive to review for a moment the ways in which systems couple, and the ways in which people couple with systems (Sheridan, 2002). Mechanical systems are nearly always coupled deterministically; that is, the mapping between input and output is governed by physical laws. If one displaces a mass in a mass-springdashpot system, the motion of the mass is predictable (to the accuracy of the measurements).

As systems become more complex, they become no less deterministic, although it may appear so. Pressing the brake pedal on your car yields a predictable set of responses, including the brake lights illuminating and the car slowing. Obviously, this is not always the response. If there is some failure in part of the system, there may be a failure to achieve the desired result. However, this is not a loss of determinism, since the brake system will always respond the same way under the same conditions of environment and system function. Rather, the extent to which pressing the brake pedal always makes the brake light illuminate and the car slow is a measure of the system's reliability, and the reliability is a function of the variability of the environment and the operation (or lack thereof) of the system components. If one knows the status of the environment and components, one can make an exact determination of the input-output mapping.

However, human behavior is not deterministic. At least to the state of our science of human behavior, even if one knows the status of environment and of the functioning of the human, a given input will not yield a predictable output. This lack of determinism is carried over in the human-machine system.

In some cases, an operator's response to an input, or that of the human-machine system as a whole, can be predicted to some degree. Such is the case with human operation within a simple feedback control system, whose response is predicted by the McRuer crossover model (McRuer & Graham, 1965). Such laws, however, are limited in application. Once the system with which the human operator is interacting becomes even modestly complex, the system's behavior will not be predictable by a simple control law.

Consider even a relatively simple machine with which humans interact, such as a microwave oven. With no goal, the number of options available to the operator is equal to the number of distinct inputs that the operator can perform. A cursory examination of my microwave suggests that this is typically a very large number, even for a simple system. Moreover, this number is an objective characteristic of the system. It does not decrease when a goal is established (e.g. “make popcorn”). However, in identifying a goal, it now becomes possible to identify a set of interactions with the machine that will result in the goal being accomplished. Note that it may be possible for there to be many sets of interactions that accomplish the goal; a procedure identifies one of those sets. A set of instructions can then be written so that the user can accomplish the procedure.

In this way the procedure acts as the coupling between the human and system. It defines the interactions the operator has with the system based on the desired goal. However, the coupling exists at two different levels. At one level is the delineation of the interactions with the system. At another level is the interaction of the human-machine system with its environment. The procedure defines the steps an operator must take to make popcorn in the microwave, but in a larger sense it defines how the microwave is to act.

In a simple system, such as the microwave example, there is little variation in how the microwave is to act across the range of possible popcorn-making procedures. In a complex, dynamic, multi-agent system, such as aviation, there is a large amount of variation in how the system acts across the range of possible procedures. The number of ways an aircraft can get a point in enroute airspace to the threshold of a runway is large, and what occurs in between that enroute point and the threshold has significant consequences for the aircraft, the pilot, the passengers, the air traffic controller, and other aircraft.

Determinism of Procedures and Checklists

It is often stated or assumed in the literature that the procedure-goal mapping is deterministic. If the procedure is accomplished, then the goal will be accomplished safely and efficiently. The focus is then fixed upon the checklist, with the implicit assumption that proper execution of a properly designed checklist will then be deterministically mapped to the goal. There is, however, a mistake with this assumption: only knowledge of the environment and functioning of the system components ensures predictability of the system's input-output mapping.

Recall that the coupling between the machine and environment, while deterministic, is not necessarily reliable. For some systems, it is possible to check the state of the system and environment and therefore predict the mapping of output from input. For many systems, however, the uncertainties in the system and the environment are such that the system is not entirely predictable, and a given input cannot be assured to have a given output.

In a simple system and with a simple goal, we can usually check the state of the environment and system to the degree necessary. If the procedure to make popcorn consisted of one step – "push the popcorn button", and we knew that the microwave worked and was plugged in to a live electrical outlet, we could be very sure that our goal would be reached when we pushed the popcorn button. However, in a complicated system, or in one in which the consequences for failure are dire, the procedure may need to implicitly ensure that the environmental and system requirements on the input-output mapping are met. Anyone who has ever jump-started a car (and even those who don't dare try) knows that it is important to ensure that the proper conditions are met before implementing the procedure.

If the proper conditions are not met (i.e. the system is in an off-nominal state), and if these conditions have not been addressed in the procedure, then the interactions with the system may need to be different than those specified by the procedure. However, while the specific interactions may have changed, and if the underlying goal has not changed as a result of the new conditions, then the procedure per se may still be valid in many respects. Procedure context, described in Chapter 3, provides a systematic way of identifying the information needed by the operator to make these determinations.

Procedures in Aviation

The experiments that will provide support for the dissertation utilize aviation tasks. Both pilots and air traffic controllers use procedures for most tasks, due to the safety critical and highly cooperative nature of the operation of aircraft in the National Airspace System. Pilots utilize checklists during each phase of flight as an aid to accomplishing all normal and emergency procedures. Air traffic procedures dictate roles and responsibilities of controllers and pilots, govern how pilots traverse various types of airspace, and dictate the transitions between phases of flight.

Of particular interest to the dissertation experiment discussed in Chapter 6 are procedures encountered at one of the most critical phases of flight – arrival into the terminal area and the subsequent final approach to the runway. In good weather, pilots can operate under Visual Flight Rules (VFR), a set of procedures designed to keep aircraft safe when they can easily see one another. In bad weather or in busy airspace, Instrument Flight Rules (IFR) must be followed.

IFR procedures are described in the Federal Aviation Regulations (FARs). The rules that govern general operating procedures are contained in FAR Part 91, while procedures for approaching airports under IFR are contained in Part 97. FAR Part 91 indicates that although all persons flying aircraft within the United States must follow the regulations, the pilot in command is the ultimate authority of the aircraft and can deviate from the rules if safety dictates (Federal Aviation Administration, 1994a).

Instrument Approach Procedures (IAPs)

Pilots flying under IFR and approaching an airport are required to fly an instrument approach procedure (IAP). These approach procedures are designed to transition a pilot from enroute airspace to the runway environment. Many approach procedures are designed so that even a minimally equipped aircraft could fly the approach.

The details for a particular approach are contained in an instrument approach plate, which identifies the navigation aid to be used, the runway to which the IAP is directed, the course to be captured and flown, and a descent profile. An example is shown in Figure 1.

In non-radar environments (or when told to by the air traffic controller), pilots must fly an IAP that takes them from any altitude and transitions them to the final approach. This procedure is designed to keep all aircraft flying approaches to the airport (and departures from the airport) separated, to keep the aircraft away from terrain obstructions, and safely and consistently turn the aircraft on to a stabilized final at an altitude that will allow a safe descent to the runway.

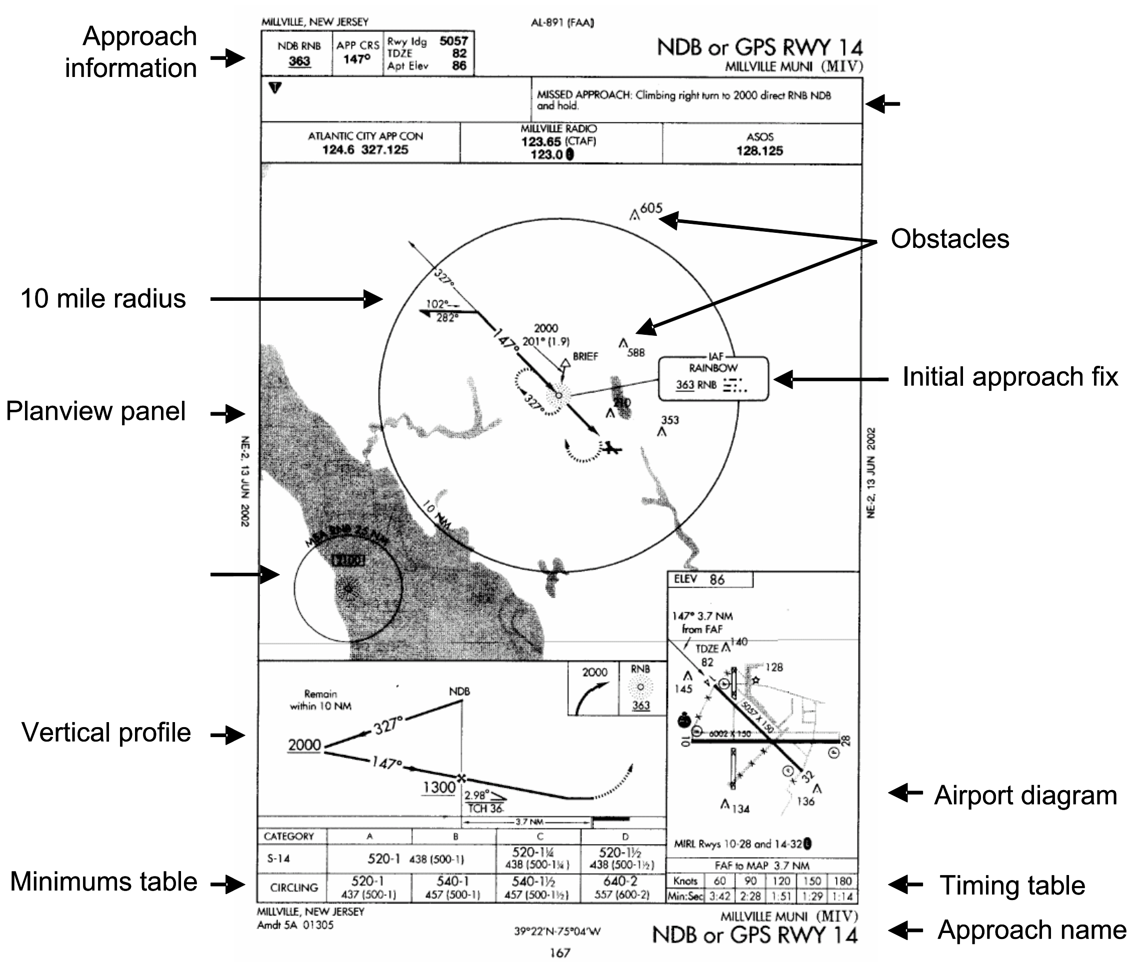


Figure 1. Instrument approach plate (Federal Aviation Administration, 2003).

To fly an IAP, pilots first cross the initial approach fix (IAF), then fly some track in order to lose altitude. Once crossing the final approach fix (FAF), pilots descend to the minimum descent altitude (MDA) (or decision height if glidepath guidance is available). Once reaching the MDA, pilots fly inbound until the runway environment is sighted or until the missed approach point (MAP), which is typically identified by either timing or by reaching a specified distance from a radio navigation aid. If the MAP is reached and a safe landing cannot be made, the pilot must "go around" – initiate a missed approach using the procedure specified on the approach plate.

There are several types of IAPs, but there are many common elements on the approach plate. Referencing the callouts in Figure 1, the elements found on all approach plates are:

- approach information, such as the identifier and frequency of the initial approach fix (IAF), the approach course, and runway information, are given in the upper left corner;
- missed approach instructions are given in the upper right corner;
- at the bottom left is a minimums table indicating the MDA and minimum weather requirements;
- a timing table, indicating how long it will take to get from the final approach fix (FAF) to the missed approach point (MAP) is shown in the lower right;
- the name and runway of the approach are found below the timing table and above the missed approach instructions;
- an airport diagram, including the elevation, is given above the timing table;
- in the center of the approach plate is a planview display of the approach, including a symbolic depiction of the approach, an indication of the IAF, indications of significant obstacles (which may be terrain or manmade), a circle around the IAF of a certain distance (in this case it is 10 nautical miles), and an indication of the minimum safe altitude (MSA – the altitude which provides 1,000 foot separation from the terrain within 25 nautical miles).

For the particular approach shown in Figure 1, the following is the information given on the approach plate:

- the IAF is the RNB non-directional beacon (NDB), its frequency is 363 KHz, the approach course inbound is 147º, the runway is 5,057 feet long, the touchdown zone elevation (TDZE) is 82 feet, and the elevation of the airport above sea level is 86 feet;
- the missed approach procedure is to turn right, climb to 2000 feet, and proceed to the RNB NDB and hold;
- for a straight in approach to runway 14 (S-14), the MDA for category A aircraft (each subsequent letter indicates a higher approach speed, category A is typically for small general aviation aircraft such as a Cessna 182) is 520 feet, the weather must be at or above a 500 foot ceiling and 1 mile visibility, and when at the MDA of 520 feet, the aircraft is 438 feet above the TDZE;
- at an approach groundspeed of 120 knots, it will take 1 minute 53 seconds to get to the MAP;
- this approach is the NDB or GPS approach (this indicates that either an NDB or a GPS is required to fly the approach) to runway 14 at Millville Muni airport (FAA identifier MIV);
- the airport diagram shows (among other things) that this airport as two runways, and that this approach is designed to land on runway 14;
- the planview display shows that this is a procedure turn (indicated by the "barb" on the depiction of the inbound course), shows the relative position of the IAF Rainbow (RNB), shows several obstacles to the left of the inbound approach

course, shows a 10 mile radius circle around RNB, and indicates that the MSA is 2,100 feet.

Procedure Turn IAP

The procedure turn IAP has several segments. The approach plate has all the common elements of any instrument approach plate (discussed earlier). There are two additional elements: the "barb" and the "remain within" distance. The barb indicates the side of the inbound course on which the course reversal maneuver is to be performed. The "remain within" distance indicates the extent within which the course reversal must be accomplished.

These elements, in conjunction with the altitude restrictions provided on the approach plate, are chosen such that an aircraft complying with those restrictions will be more than 1,000 feet above any terrain feature until after the FAF.

The FARs state only that the procedure turn depiction must show "the outbound course, direction of turn, distance within which the turn must be completed, and minimum altitude (Federal Aviation Administration, 1994b, section 97.3)." This means that, for a procedure turn depicted with a barb, the point at which the turn inbound is made, and the type and rate of the turn, is at the pilot's discretion. Typical maneuvers taught (and commonly used) are the 45º-180º maneuver, the racetrack pattern, the tear-drop, or the 80º-260º maneuver. These maneuvers are shown in Figures 2a through 2d. The headings indicated on the approach plate correspond to the 45º maneuver, but this maneuver is not required. Some depictions of a procedure turn indicate a particular ground track, and these are to be flown as depicted.

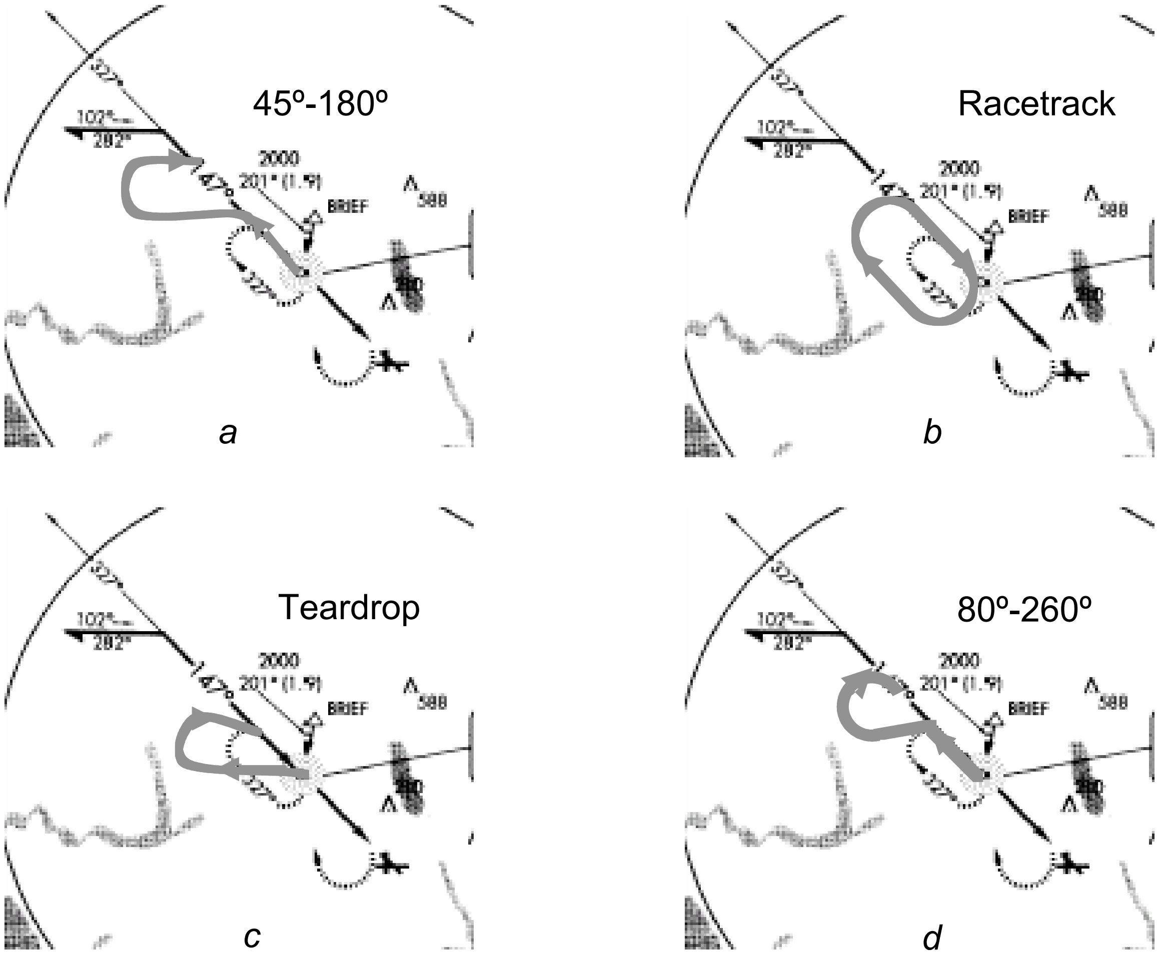


Figure 2. Entry maneuvers for procedure turn IAP.

The 45º-180º maneuver consists of a turn to the inbound course outbound for approximately 45 seconds to one minute, then a 45º turn towards the protected side (remaining on that heading for approximately 1 minute), then a 180º turn in the opposite direction as the last turn. This 180º heading should give the aircraft a 45º intercept heading to the inbound course. The aircraft then intercepts and follows the inbound course (and the remainder of the procedure).

The racetrack pattern is a technique typically used when the aircraft is already flying close to the inbound course to the procedure turn fix. In this case it would take a 180º turn to start a 45º maneuver, and the aircraft would be well off of the inbound course outbound by the time such a turn is completed. Instead, the aircraft simply turns 180º towards the protected side and flies that course outbound for one to one and a half minutes. The aircraft then turns 180º again (in the same direction), which should put the aircraft back on the inbound course.

The tear-drop maneuver consists of turning immediately to the 45º heading depicted on the approach plate once passing over the procedure turn fix, and flying that heading for one to one and a half minutes, then turning 135º back towards the inbound course. These turns should result in the aircraft being close to on course inbound.

The 80º-260º maneuver is similar to the 45º maneuver except that the 80º heading is not maintained. Once the heading is obtained, the turn back towards the inbound course is commenced. This maneuver should result in the aircraft being close to on course inbound.

The "remain within" distance must be complied with for terrain avoidance reasons. Outside this region 1,000 foot obstacle clearance is not guaranteed if the aircraft is below the MDA. The standard maneuvers described above have been shown to remain within 10 miles. If the pilot uses any other maneuvers or if the "remain within" distance is 5 miles, then some planning is required, and typically different times for when to turn need to be used in order to stay within the protected airspace.

Chapter 3

Supporting Literature

Procedures

Due to the safety critical nature of aviation, the environment is highly proceduralized, as can be found in other safety critical domains (such as nuclear power). Yet the design of checklists and procedures for aviation has not been directly investigated until recently (Degani & Weiner, 1994), and to this day is not the subject of a great deal of research effort. The reason is probably best reflected in a response early investigators got to their inquires – "checklists, they are simple and straightforward, so what is there to study about them (Degani & Weiner, 1990, p. 2)?"

Although several aviation accidents related to procedures occurred earlier, the first real research interest in procedures came about outside of the aviation industry. On March 28, 1979, a minor malfunction at the Three Mile Island (TMI) nuclear facility spiraled dangerously out of control due in large part to poor interface and procedure design. The malfunction was not responsible for the severity of the accident – the resulting actions by the well-trained operators were. A nuclear power plant is extremely complex, and the operation of the plant, controlled through procedures, displays, and automation, is quite distal from its functioning. For this reason, those operating procedures, displays, and automation are critical to the plant's functioning. The TMI accident demonstrated that in a system of critical safety and high complexity, the value of good procedures and good displays can outweigh good technical knowledge of the underlying functioning of a system, if that knowledge goes unsupported.

As a result of this accident, a flurry of research activity regarding displays and procedures for nuclear (and similar) plants commenced. This research, and the attention the subject received, did not stem the tide of procedural failings, however. Several groups of investigators found that failures with procedures remained the cause of the majority of human-factors related accidents (Marsden, 1996).

The first direct research on aviation checklists and procedures was motivated by a string of accidents related to checklists. In 1969 a Pan American World Airways Boeing 707 attempted a takeoff without having deployed flaps and crashed, prompting the safety board to call for a review of cockpit checklists (National Transportation Safety Board, 1969). That recommendation was not heeded. The following decades witnessed numerous accidents with checklist or procedural factors involved in the accident description. The ways in which checklists can contribute to accidents are numerous, and not always listed as an official causative or contributory factor. NTSB reports frequently mention some inadequacy regarding procedures or procedure-following. Following is a list of quotations from the NTSB online database (National Transportation Safety Board, 2004) that demonstrate the range of checklist problems (abbreviations replaced with words in parentheses):

- " … the (flight crew) failed to follow procedural checklist requirements & to detect & correct a mistrimmed stabilizer before the (aircraft) became uncontrollable … (National Transportation Safety Board, 1983a)"
- " … the (captain) became disorientated (sic) and inadvertently reentered (runway) 25R … A DC-10 had already begun its takeoff roll (and) rotated and began its climb … (crossing) over … (Flight) 1605 about 200 (feet)

A.G.L. The first officer (reported) that while the (captain) was taxiing … he was busy completing the After (Landing) Checklist. (National Transportation Safety Board, 1988a)"

- "FAA approved procedural checklist had omitted (a) critical step in (the) manual gear extension procedure. (National Transportation Safety Board, 1988b)"
- "A second (alternate) gear extension procedure … is not published in the crew manuals, taught during initial training, or available on any checklist. (National Transportation Safety Board, 1990)"
- "Numerous genuine and 'phantom' TCAS traffic alerts were received during the approach, which distracted (the crew) from their checklist. (National Transportation Safety Board, 1994a)"
- "The evidence indicated that the flaps and slats were in the up/retract position and had not been deployed for (takeoff). Neither (pilot) recited the items of the taxi (checklist). (National Transportation Safety Board, 1987a)"
- "The crew failed to assure that the RPM levers were full forward in the takeoff (100 percent) position … Both crew members had relatively little (flight) experience in the BAE-3101 which differs in RPM control lever procedures from other (aircraft) … (National Transportation Safety Board, 1987b)"
- "… the (flight crew attempted) to takeoff without the wing flaps and slats properly configured for takeoff … Contributing to the accident was

Delta's slow implementation of necessary modifications to its operating procedures, manual's (sic), checklists … (National Transportation Safety Board, 1988c)"

- "The rudder trim position was found to be 16 (degrees), or full, left deflection of rudder. The (flight data recorder) recorded this position from engine start at the gate. (National Transportation Safety Board 1989)"

The last four quotations were taken from accidents in which there were a total of 172 fatalities, 33 serious injuries, 83 minor injuries, and 58 uninjured, and they happened over the course of 28 months - from May 1987 to September 1989. It was, more than any other incidents, these that caused the focus on aircraft procedures and checklists. There was, at the time, almost no research on how checklists should be designed (Degani & Weiner, 1990) with the notable exception of several studies examining the use of electronic checklists (Rouse & Rouse, 1980) (Rouse, Rouse, & Hammer, 1982).

The initial concentration of effort was on the paper checklists itself, evaluated as a display aid. However, it became apparent that the problems were much more complex. Procedures and checklists are part of the "operating concept" (Degani & Weiner, 1994) of the system, and are crucial to its successful and safe operation. Among the issues analyzed over the past decade concerning procedures are: general limitations/deficiencies of paper checklists, social and organizational issues, interruptions and distractions, underreliance and overreliance, procedure context, and automating checklists and procedures.

Limitations/deficiencies of checklists

The paper checklist was designed to be a memory aid to ensure that certain parts of critical procedures were accomplished. There are plenty of examples of incomplete, inaccurate, poorly organized, and poorly written checklists. In addition, however, are (Gross, 1995):

- Checklists can be hard to locate or hard to access.
- Checklists may branch, creating a navigation problem.
- Checklists can be different or even disagree with flight manuals and training.
- Checklists vary significantly between equipment and company.
- Emergency checklists are often organized very differently by different organizations.
- Checklists are sometimes unclear or contain “training-type” information.

This is but one group of classifications. There are others, but they can be categorized as having to do with either design of the procedure or implementation of the procedure.

Social and organizational issues

Procedures are not self-contained entities. They exist in context, and part of that context is the individuals who interact with the procedures, and the organizations that create and implement the procedures. Issues of philosophy, policy, implementation, and attitudes all affect the utility of procedures.

Most organizations working with high-risk systems have a strong belief in procedures. The aviation and nuclear industries are two prime examples. Both create standard operating procedures that cover a substantial portion of normal operations, and nearly all emergency operations. These operating procedures often reflect the philosophy of the organization (or at least what is currently emphasized in the organization). As much as possible, it is suggested that the design of procedures reflect this philosophy and the policies of the organization that flow from that philosophy (Degani & Weiner, 1994).

Of course, different philosophies and policies lead to different procedures. Checklists for the same aircraft often differ substantially between companies (Degani & Weiner, 1994). This is true across equipment as well as across companies. The result is that even within a company, procedures will differ as pilots transition across different airframes.

This is also true in many other, more mobile, settings. There are procedures in most work settings, and associated with many pieces of equipment, including consumer electronics (and in putting together new furniture or toys). These procedures will vary significantly across similar applications, creating difficulty and in some cases error.

In multi-person systems, procedures are adapted by different members of the organization. Members utilize procedures in the context of their work, and adapt them based on their own experience and skill (Hughes, 1992).

Operators also adapt procedures due to the uncertainty present in the environment. An example of this is a case of deliberate nonconformance to procedure:

> Descending (through) 15000 (feet) into Nassaue the #2 (engine) was shut down due to low oil (pressure) at 16000 (feet). Returning to Miami the #3 (engine) flamed out, (and) 3 (minutes) later the #1 (engine) flamed out. The (aircraft) began descending without power from 13000 (feet). At about 10000 (feet) the flightcrew announced that ditching was imminent. The #2 engine was restarted … and the (aircraft) made a (one-engine) landing at Miami. All o-ring seals in the master chip detector (assemblies) in the (engine) lubrication

> system were missing causing oil leaks in all (engines). Proper procedures to remove, reinstall, and inspect the detectors for oil leaks were available. The foreman knew that mechanics were not routinely replacing o-ring seals. Accident was 9th chip detector occurrence since procedures were revised 12/81. (National Transportation Safety Board, 1983b)

Several issues were at play in this near-accident. New procedures had recently been developed, and mechanics were apparently unable to follow them. The pilots disregarded procedures by not shutting down engines with low oil pressure lights on (the two engines that later flamed out - #3 and #1), because they believed that the chances of three engines having low oil pressure simultaneously was “one in millions I would think” (Norman, 1988). This calculation was based on, in part, the assumption that every one else had followed procedures. Yet their actions in violating procedures were correct in the sense that the procedure would not have helped them. The L-1011 cannot operate (well) without engines, and they would, at some point, have been required to create their own procedure for dealing with something that went well beyond the scope of what procedures they had.

The previous example points to a pervasive feature of the proceduralization of the work environment. The basic framework of procedures provides a great deal of information about the actions of others. They provide general predictive information about the environment (including other operators, vehicles, etc.), and without this information, operators can only react to events as they occur with no foresight (Pritchett & Yankosky, 2000).

As a result of these variations, some procedure guidelines for airlines were developed, relating to (amongst other things) social and organizational issues (Degani & Weiner, 1994):

- A feedback loop from line pilots to flight management and procedure designers should be established. This feedback loop should be a formal process. It must be maintained as a non-punitive, reactive system, with mandatory feedback from management to the initiating line pilot about the progress of his or her report and/or suggestion.
- It is essential that management develop a philosophy of its operations. This is particularly important for operating automated cockpits.
- Management, through the use of the feedback loop and the line check airman program, should be watchful of techniques that are used on the line. Techniques that conform to procedures and policies should not be interfered with. Techniques that have a potential for policy and procedure deviation should be addressed through the normal quality assurance processes. Techniques that yield better and safer ways of doing a task may be considered for SOP.

Interruptions and distractions

A frequent problem with procedures and checklists is distractions, both from the procedure and by the procedure. The Aviation Survey Reporting System (ASRS) is replete with examples of flight crew being distracted from a procedure, resulting in skipped steps or worse (Monan, 1979). In other cases, the checklist or procedure itself is a distraction. These instances are not entirely a problem with the procedure or checklist, however, but relate more to the general workload, poor planning, or unfortunate timing of events.

One method of avoiding missed steps due to interruptions is the use of some type of “placeholder”. This may be in the form of verbalization (Linde & Goguen, 1987), or some type of pointer such as in scroll checklists (Degani & Weiner, 1990). These methods are infrequently used, however, and scroll checklists are generally not used for in-flight checklists.

Underreliance and overreliance

As mentioned previously, nonconformance to a procedure can be unintentional or intentional. The latter case is sometimes due to “underreliance” on the procedure. This term is frequently applied to automation (Kirlik, 1993) when the automation is regarded as a more complicated means of accomplishing a task. In this case, operators will often not use the automation. The same is true of procedures. If the procedure is seen as overly burdensome, unreliable, or unnecessarily complex, operators may not use, or may find ways around, the procedure.

The opposite case, overreliance, also has an automation counterpart (Parasuraman & Riley, 1997), where operators will use the automation even when it is malfunctioning or its use is inappropriate. Procedures in some ways may be more prone to this type of failure. Four ways in which overreliance can become a factor have been elicited by researchers (Ockerman & Pritchett, 2000):

- The procedure may be accurate but its intention may be misinterpreted.
- The procedure may be accurate and its intention may be interpreted properly, but it may be used outside its intended range of application.

- The operator may use an inappropriate strategy in implementing the procedure.
- Parts or all of the procedure may be wrong.

Overreliance can be mitigated in a number of ways, including using design guidelines for procedures and procedure support such as those proposed in this dissertation. The operator should be thoroughly trained on the system and the procedures, including an understanding of the assumptions of the designer. A designer may embed “additional” information that is not specifically part of the procedure, which may include having some type of automated assistance in accomplishing the procedure. In addition, the procedure can be designed to better situate the operator in the context of the procedure. Lastly, a rigorous design and testing process can hopefully eliminate errors in the procedure.

Deliberate Noncompliance

Previous research efforts have viewed deliberate noncompliance as due to individualism, complacency, frustration, technique, or even humor (Degani & Weiner, 1994). While these are certainly true, they only represent an undesirable class of noncompliance. Yet despite the best intentions of procedure designers, operators can find themselves in situations in which procedures are wrong or don’t exist at all. In these circumstances the operator may choose to violate procedure.

In doing so, the operators must create their own procedure, or modify an existing one. In July 1989 a United Airlines DC-10 had a failure which destroyed the #2 engine and one of three hydraulic systems, and which subsequently severed the lines of the two remaining hydraulic systems. The control surfaces of the aircraft had no manual backup

– they could only be controlled by hydraulics. The loss of all three hydraulic systems had been thought impossible by designers, and crews had never trained or heard of any procedures to handle the aircraft in that situation. Nonetheless, the aircrew was in that situation and had to develop a procedure. The pilot later recalled,

> And as a result (of having no hydraulic systems), we had no ailerons to bank the airplane, we had no rudder to turn it it, no elevators to control the pitch, we had no leading-edge flaps or slats to slow the airplane down, no trailing-edge flaps for landing, we had no spoilers on the wing, to help us get down, or help us slow down, once we were on the ground. And on the ground, we had no steering, nose wheel or tail, and no brakes. So what we had … was the throttles on #1 and #3 engine to control us. And by manipulating those throttles, we were able to somewhat control the heading, by skidding the airplane into a turn. And controlling the pitch was just about out of the question (Haynes, 1991).

The pilots created a procedure to control the aircraft without hydraulics and performed a somewhat controlled crash landing in Sioux City, Iowa. The pilots' ability to not rely on a predefined procedure saved the lives of over 180 people on this flight.

Automation for Procedures

Some previous efforts have been made to use automation to support procedure following. In a study designed to test automated procedure aiding in an industrial setting, an automation tool that could control the system was modified to provide procedure guidance to operators in an attempt to enhance performance. Surprisingly, operators still performed worse than the automation tool, and were found to use different courses of action in dealing with plant operation. It was found that the automation had been programmed with heuristics designed to maximize production, a goal that was not reflected in the operational procedures (Knaueper & Morris, 1984).

Automation of checklists and procedures was initially seen as a way of guarding against error, with the automation monitoring implementation of the procedure. Additional benefits, such as providing a larger set of procedures or even automating creation of procedures "on-the-fly" were also investigated (Hammer, 1984). Some evaluations of automating procedures have produced positive results (Rouse & Rouse, 1980), while others have produced mixed or even negative results (Rouse, Rouse, & Hammer, 1982)(Converse, 1994). In addition, one study found that pilots may have overreliance problems on automated checklists (Palmer & Degani, 1991).

In each of these cases, the aids were designed to assist in essentially following checklists, and assumed universal compliance. The aids assisted operators in following the steps of the procedure, but did not consider the overall context of the procedure.

Electronic Flight Bag

A recent development has been the introduction of "electronic flight bags" (EFBs), which are essentially a laptop or other small portable computer that replaces all of the paper charts that pilots normally carry with them. Although not approved as a replacement for paper charts by the FAA, some aircraft have these installed and they are being used in addition to paper charts. EFBs have the potential to not only replace paper charts and operating manuals, but also could be used as electronic checklists, could receive real-time weather, and could even interact with flight systems (Shamo, Dror, & Degani, 1998). Since the definition of standards for these devices have only just been developed (Chandra, Mangold, & Riley, 2002), certification efforts have lagged behind development and deployment.

Measures

Situation Awareness

As more complex automation is added to flight decks and air traffic facilities, and as more information is provided to operators, accessing and comprehending that automation and information becomes problematic. Researchers have begun to understand that not only do they need to know how to display information, but also how operators represent and utilize that information, in order to provide a better match between technology and the human (Endsley, 2000).

Situation awareness (SA) has been a particularly important concept in aviation, since safe and successful operation in aviation requires a great deal of knowledge about the environment outside of an individual aircraft (Harwood, Barnett, & Wickens, 1988). Many aviation accidents can be attributed to a lack of knowledge of the environment external to the aircraft (collisions with aircraft and the ground, for example) (Bolman, 1979). For air traffic controllers, their knowledge of the "big picture" includes not only the states of the aircraft in their own sector, but those of neighboring sectors, airports, weather, and much more. Attempts to understand how pilots and controllers obtain and maintain this "big picture" have been the focus of SA research.

Military aviators, since as far back as the first World War, have understood the importance of SA, and instructed their pilots in developing and maintaining good SA. One of Germany's top World War I's fighter aces, Oswald Boelcke, listed among his "dicta" that "the pilot must acquire the habit of 'taking in' unconsciously the general progress of the whole mult-aircraft dogfight going on around the individual combat in which the pilot will become involved … (so that) no time (is) wasted in assessment of the general situation after the end of an individual combat (Hacker, 1984, section IV, number 6)." Boelcke also prescribes knowledge of one's own machine, the enemy's machine, and navigational fixes. SA has been a significant aeronautical training topic since that time.

As a research topic, however, the concept was mostly ignored by researchers until the 1980s. Due to increasing flight deck and air traffic automation, pilots' and air traffic controllers' role as supervisor of these systems was increasing, reducing the time they could spend in developing SA. At the same time a great deal of new sensor information was becoming available to designers of aviation automation, information that could be used to reinforce the controllers' and pilots' SA. Researchers began looking into what SA is, what affects an operator's ability to construct SA and to keep it, and how it might be measured.

Situation awareness research

Many definitions of SA exist, with most agreeing that SA is, at least in part, the comprehension of elements of the environment that have (or may have) some bearing on the task being accomplished.

Several efforts at defining SA have been put forth. Some efforts have viewed SA as a static, information-driven product, some have viewed it as a dynamic process, while others have viewed SA as a high-level description of certain aspects of task behavior. One of the most widely quoted definitions of SA is: " … the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1999, p. 258)." Another researcher has called SA "… an integrated understanding of factors that will contribute to the safe flying of the aircraft under normal or non-normal conditions (Regal, Rogers, & Boucek, 1989, p. 65)," These definitions describe SA as a product, something that an operator either has or does not have. Typically these authors, although acknowledging its importance, defer the production of SA as a separate topic of inquiry. The argument is that SA-related errors are caused by the lack of SA, which is then only in turn affected by the lack of production of SA. So comprehending the SA product can help us understand why errors were made, whereas understanding the process can only help us understand why SA is lacking.

SA has also been described as a process, such as in referring to SA as "(an) adaptive, externally-directed consciousness" (Smith & Hancock, 1995, p. 137), and "the integration of knowledge resulting from recurrent situation assessments (Sarter & Woods, 1991, p. 45)." Some researchers, as mentioned above, consider the process of obtaining SA separate from the product, and refer to this process using this term – "situation assessment". The process-based SA research has mostly gone towards understanding how SA is obtained and maintained.

Some researchers have called into question the utility of the SA concept. Many feel it is a high-level concept that does not have sufficient granularity to really explain anything. SA has been referred to as simply a description of observations of humans operating complex systems within a dynamic environment (Billings, 1995) and as simply equivalent to expertise (Crane, 1992). As such, SA is a description of a set of cognitive processes that are used together, and is really only useful in categorizing or grouping behaviors and errors.

Researchers have also examined how SA is obtained and maintained. Here the sources of the information (visual, auditory, tactile, other sensory input, knowledge of procedures and regulations, etc.) are diverse and often clearly identifiable. The operator's ability to detect, comprehend, synthesize, and project this information is generally of more interest to researchers. It is interesting to note that these abilities begin with the most fundamental cognitive processes (detection), and progress to very sophisticated concepts (projection), and that they are typically studied separately, often by different groups of researchers.

Detection of information involves the well-studied (if not well-understood) processes of perception and attention. Researchers have studied how these processes influence SA, including how the limits of working memory affect SA, how long term memory may be used in SA, and how task objectives influence attention.

Comprehension of information for use in SA has also been studied. Several researchers have examined how operators may use recognition or pattern-matching to develop SA, although there is also evidence that this strategy is not sufficient to explain all SA. Expectations also influence the development of SA, both positively and negatively.

Measuring SA

SA, as a set of knowledge, has been difficult to measure. Typical measures can be broken into three categories: explicit (or knowledge-based), implicit (or performance-based), and subjective (Vidulich, 1992). Examples of explicit methods include participants' recall of a situation (post-hoc), an ongoing narrative provided concurrently with the task, or freezing a scenario and questioning the participant about the decisions, events, or the task environment. These measures can be compared to actual state of the system to provide a better measure, but have been criticized as too subjective (in the case of recall) (Fracker, 1991), or too intrusive (in the case of freezing a scenario) (Sarter & Woods, 1995).

Implicit measures examine task performance, and correlate that with SA. These measures are generally unobtrusive and objective, and (as mentioned earlier) can be used in conjunction with explicit measures. It has been suggested that these measures can succeed where explicit methods cannot, particularly in situations where a determination of the timing of events is important and where the subject may be unaware of his or her deficiency (Pritchett, Hansman, &Johnson, 1995). These measures may examine performance at the task overall, or alter the task to determine whether the subject notices the change.

Subjective ratings are assessments of SA made either by an observer or by the subject. These ratings can use a number of different scales, and can be either direct or relative (such as by comparing SA in one situation to SA in another situation). Although these ratings can be affected by a number of factors, including task performance, one technique that has been studied extensively is the Situation Awareness Rating Technique (SART) (Taylor, 1989). This technique has the operator rate 10 constructs on a seven point scale. The constructs are grouped into three categories: attentional demand (which includes the instability, variability, and complexity of the situation), attentional supply (including arousal, spare mental capacity, concentration, and division of attention), and understanding (which includes information quality and quantity, and familiarity). SART has been found to be more sensitive than overall subjective measures of SA (Selcon & Taylor, 1989), although some researchers feel that SART confounds SA with workload (Jones & Endsley, 2000).

NASA TLX

The NASA Task Load Index (TLX) workload rating system was adapted from an earlier NASA Bipolar Rating Scale (Hart & Staveland, 1988). There are six dimensions to the NASA TLX, three related to the subject (mental, physical, and temporal) and three to the interaction of the subject with the task (effort, frustration, and performance). TLX has been widely used in a variety of domains to measure workload.

TLX users rate the six dimensions on a 20 point scale, and perform a pairwise comparison on those dimensions. The pairwise comparisons are then used to weight the user's workload ratings.

Procedure Context

One method of mitigating over- and underreliance is through presenting procedure context (Ockerman, 2000). In an aircraft inspection task, overreliance on the checklist was reduced when pilots were presented with procedure context information (Ockerman & Pritchett, 1998).

As discussed by Ockerman (Ockerman, 2000), a great deal of information that could be useful for understanding the underlying assumptions and constraints of a particular procedure was considered by the developers of that procedure. This "context" of the procedure – its intentions, assumptions, ordinality, etc. is often lost once the procedure is implemented. Knowledge of these elements gives pilots access to a great deal of information about a procedure.

The context of a procedure provides the user with information that can enable him or her to make informed judgments concerning implementation of steps in the procedure, deviations or alterations to the procedures, and appropriate levels of reliance on the procedure. Procedure context consists of two categories – explanatory and locational, and their elements.

Explanatory context provides background information on the procedure, indicating purposes and interrelationships within the procedure. It helps the user apply a strategy to accomplish the procedure, and aid him or her in understanding consequences of not complying with the procedure, or in enabling him or her to safely alter the procedure during execution. Normally much of this information is not transferred by the procedure designers and is lost. If retained, however, the information can be provided to the user through training, documentation, or displays. Elements of explanatory context are:

- Intention – this reflects the overall goal state of the procedure.
- Rationale – this reflects the reasons for individual steps.
- Boundary conditions – this reflect the conditions under which the procedure is assumed to be operating.
- Triggering conditions – this reflects external conditions that may cause a procedure to begin, branch, or end.
- Temporal construct – this reflects the time window in which the procedure is assumed (or is required) to be accomplished.

- Ordinality – this reflects the requirements for the order in which steps must be accomplished.
- Necessity – this reflects the degree of requirement that the step be accomplished.
- Reversibility – this reflects the degree to which actions accomplished as part of the procedure can be "taken back".
- Appropriate specificity – this reflects the degree to which the procedure captures the detail of what needs to be accomplished.

Locational procedure context is intended to provide information concerning the physical ordering of the procedure. Its elements are:

- Previous actions – this relates to the actions that have been already accomplished.
- Following actions – this relates to the actions that are upcoming.
- Location indication – this relates to where in the global procedure the current step resides.
- Forking – this relates to how a procedure might branch.

Chapter 4

Preliminary Experiment Investigating Pilot Use of Procedure Information In A Parallel Approach Task

This chapter describes a preliminary experiment which investigated pilot use of procedure information during instrument approaches (Landry & Pritchett, 2002a; Landry & Pritchett, 2002b). In this experiment these roles of procedures were investigated by examining pilot compliance during a closely spaced parallel approach procedure, their reaction to another aircraft's noncompliance to procedure, and their use of displays that utilize procedural information.

The results of this experiment movitated the dissertation research by demonstrating that:

- pilots used a heuristic strategy to conduct the procedure,
- compliance to procedure cannot be assumed,
- noncompliance to procedure may be a necessary adaptation to the task environment, and
- pilots were unable to interpret the consequences of the other aircraft's noncompliance and will continue to use procedure information in situations in which the procedure no longer applies.

Background

The proposed approach procedure, "paired approaches", places two aircraft on instrument approaches to closely spaced parallel runways, with one aircraft initially offset behind the other. Currently such approaches are only allowed in visual conditions or when closely monitored by an air traffic controller with a fast update radar and dedicated display (the system is called the "precision runway monitor"). This new instrument procedure is being investigated to attempt to increase airport throughput during bad weather and to runways closer than allowed with the precision runway monitor system.

The trail aircraft maintains a position relative to the lead aircraft (Stone, 1996; Pritchett, 1999; Hammer, 1999). Such a position (called the "safe zone") guarantees that neither aircraft will be in danger of loss of separation within a certain time window should the other depart its approach path, and that neither aircraft will be affected by the other's wake.

Two different underlying bases can be used to determine the safe zone (Pritchett & Landry, 2001). The first uses procedural information; i.e. the "predicted" safe zone is calculated assuming that the aircraft are following the pre-specified approach procedure. The predicted safe zone provides a spatial boundary which is small and unchanging throughout the approach, but which is inaccurate if either aircraft is violating the approach procedure.

The second basis for the safe zone uses real-time information; i.e. the "actual" safe zone is based on the current states of both aircraft. The actual safe zone is continually updated, and presents a spatial boundary which is as large as possible for the immediate context, but constantly changing in size and location.

This preliminary study examined the relative merits of the different underlying conceptual bases of the safe zone, which imply different control strategies and suggests different types of monitoring for unusual situations. In particular, since the predicted safe zone is not valid if either aircraft deviates from procedure, it was expected that pilots would more closely monitor compliance when using the predicted safe zone. In addition, pilot conformance to procedure and reaction to the other aircraft's nonconformance was investigated.

Method

Participating pilots (12 male airline pilots current or previously qualified in glass cockpit aircraft) were asked to fly approaches using Georgia Tech's Reconfigurable Flight Simulator (RFS) (Ippolito & Pritchett, 2000). The RFS is a medium fidelity simulator which ran on a Pentium III desktop computer. The experiment setup is shown in Figure 3. One aircraft was simulated and flew a present approach, while a second aircraft was flown by the participant pilot. The participant pilot's aircraft was programmed with a dynamic model and cockpit systems representing a Boeing 747-400. The instruments are shown in Figure 4.

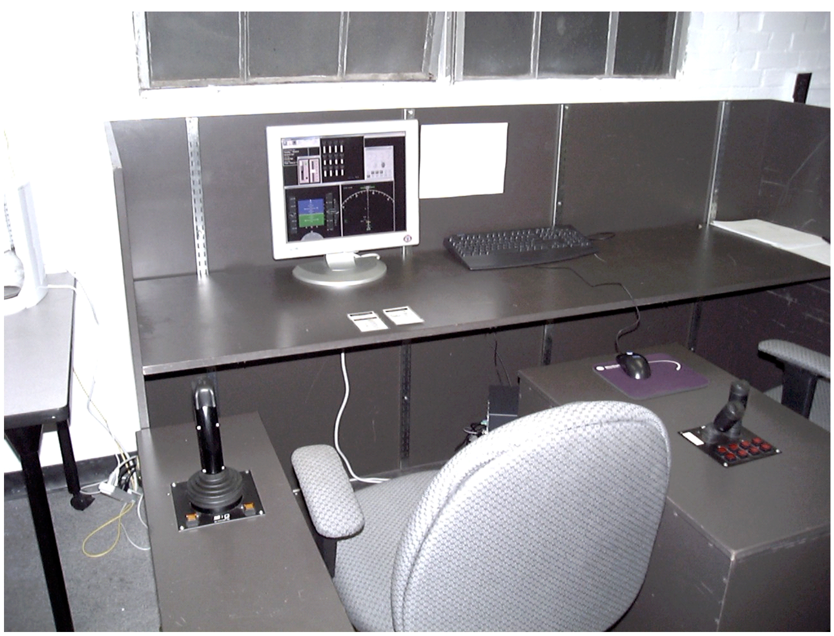


Figure 3. Experiment setup for preliminary experiment.

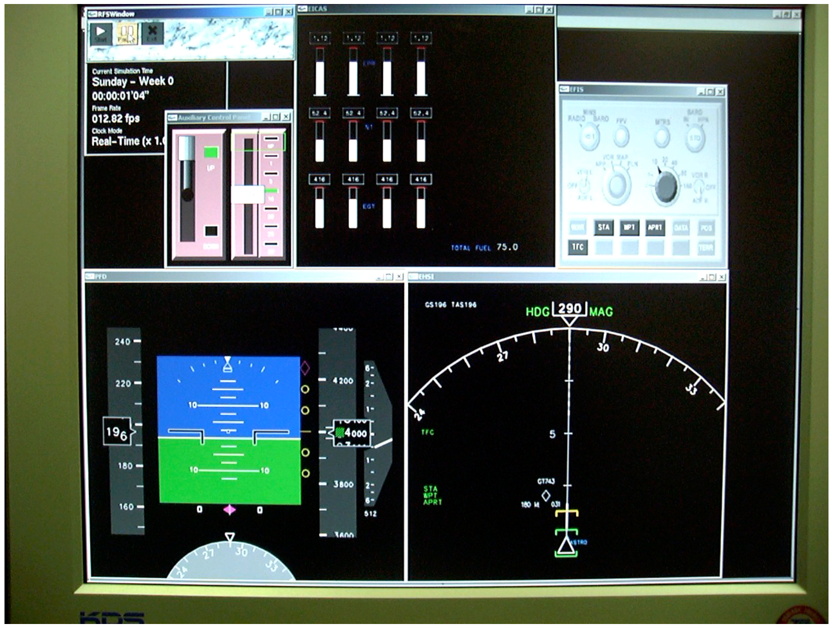


Figure 4. Instruments for preliminary experiment.

The navigation display (ND) included an overlay of traffic information about the aircraft on the other approach and the safe zone presentations, which were displayed as staple shaped brackets (Figure 5). The inner (green) brackets in Figure 5 represented the predicted safe zone, while the outer (yellow) brackets represented the actual safe zone.

Pilots were given detailed briefings on the simulator and the procedure, and given an opportunity to practice with each until they felt comfortable. In their briefing on the safe zone, it was stressed that a position within the actual safe zone was safe for the next 30 seconds from collision and wake turbulence regardless of the actions of either aircraft, while a position within the predicted safe zone was similarly safe, but only if the two aircraft were not deviating from the approach procedure.

Figure 5. Navigation display for preliminary experiment showing both safe zones.

For either safe zone indication, if a safe zone limit was exceeded, the protection provided was no longer guaranteed, and it was recommended that a missed approach be executed. The missed approach procedures were provided on the approach plate, and indicated both a climb and a turn away from the other approach path.

The participants were instructed to fly an instrument landing system approach, while remaining within the safe zone. The participant pilots flew the trail aircraft, with the lead aircraft being the scripted simulated aircraft. Each run began at approximately 20 miles from runway threshold on the localizer and at approximately 200 knots true air speed (KTAS). The participants were instructed that ATC had told them (and the lead aircraft) to maintain 180 KTAS, plus or minus 10 knots, until 5 miles from runway threshold, where they could slow to their normal approach speed of 148 KTAS.

Each participant pilot flew 10 data collection runs. The first nine runs represented a two factor design with three safe zone displays and three blunder types. The three displays refer to the conceptual basis of the safe zone, as follows:

- Predicted safe zone display: The predicted safe zone was shown on the ND.
- Actual safe zone display: The actual safe zone was shown on the ND.
- Both safe zones display: Both safe zones were shown on the ND, allowing the pilot to directly compare the two types of information. In this case the pilots were briefed that they could exceed the predicted safe zone limits as long as they remained within the actual safe zone limits.

The blunder type refers to the type of noncompliance committed by the lead aircraft:

- No noncompliance: a baseline in which the lead aircraft complied with all procedural restrictions.
- Speed noncompliance: The lead aircraft slowed substantially below the approach procedure’s minimum allowed speed, as if this aircraft were configuring and attaining final approach speed 5-10 miles before allowed by approach procedures.
- Lateral noncompliance: The lead aircraft turned toward and crossed the participant’s approach path, in the form of a turn to a new heading commonly used as a blunder model.

Once the participant completed these nine runs, he (all participants were male) flew a tenth run with one of the three safe zone displays in a combined deviation scenario: specifically, the lead aircraft first slowed below the minimum allowed procedural speed, and then the lead aircraft also turned toward and crossed the trail aircraft's approach path. The complete experiment matrix is shown in Table 1.

Table 1. Experiment matrix for preliminary experiment.

Display of Safe Zone

	PROCEDURAL	REAL-TIME	BOTH
No noncompliance	A	B	C
Speed noncompliance	D	E	F
Lateral noncompliance	G	H	I
Both speed and lateral	J1	J2	J3

In addition to observations of pilot behavior and comments, the following quantitative measurements were taken:

- Pilot detection of non-procedural actions by the lead aircraft, and the ability of the pilot to describe what the lead aircraft did;
- Approach stability as indicated by the variation about mean in the throttle movements, elevator movements, and glideslope tracking

Results

The results of the participants' detection of lead aircraft noncompliance is shown in Figure 6. Subjects failed to detect 4% of lead noncompliance, and mis-detected 18% of the lead noncompliance actions.

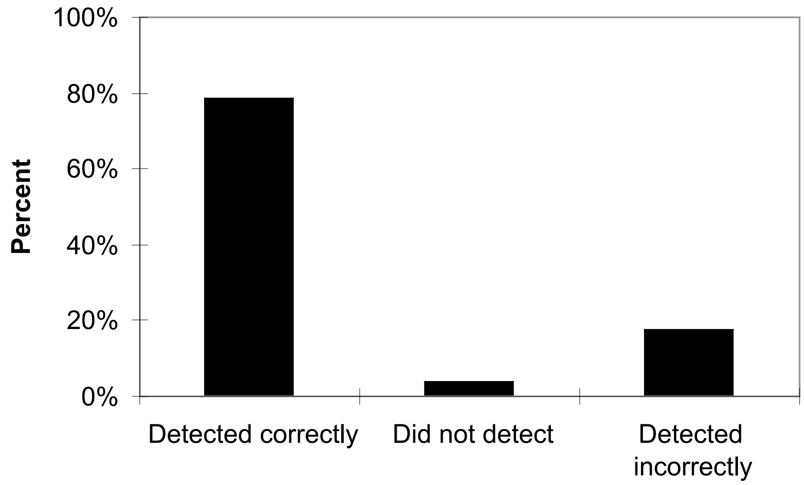


Figure 6. Pilot detection of lead aircraft noncompliance.

The detection results were analyzed with respect to the independent variables, as shown in Figure 7. Kruskal-Wallis tests indicated no significant results with respect to either independent variable.

The standard deviations of throttle movements, elevator movements, and glideslope deviation, indicative of the stability of the approach, were also analyzed. The least-square means of the standard deviations with respect to the independent variables are shown in Figure 8. Elevator movement standard deviations were normally distributed; glideslope and throttle movements were transformed by taking the square root.

The differences in means were significant between levels of the blunder variable for the square root of throttle standard deviation only. An ANOVA for the square root of throttle standard deviation indicates that there is a significant difference ($p<0.001$), and Tukey simultaneous tests indicate that the differences were all significant ($p<0.02$) except between the both blunder scenario and the two single blunder scenarios (lateral and speed).

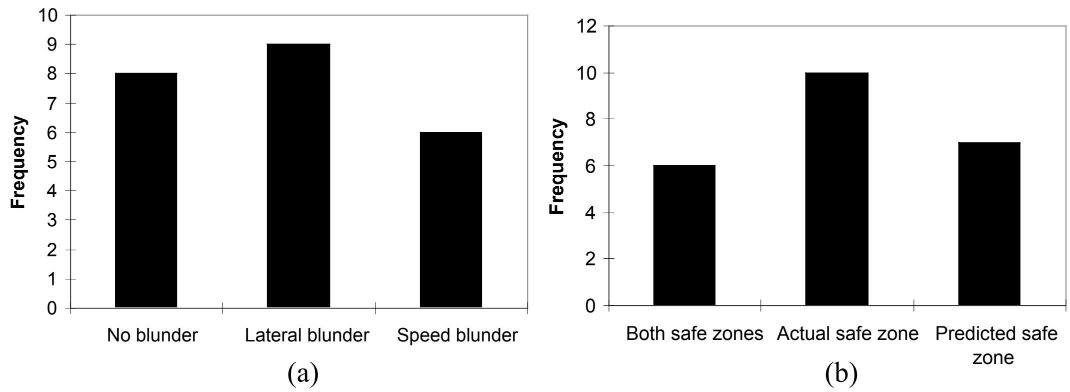


Figure 7. Pilot errors in detection by blunder and display.

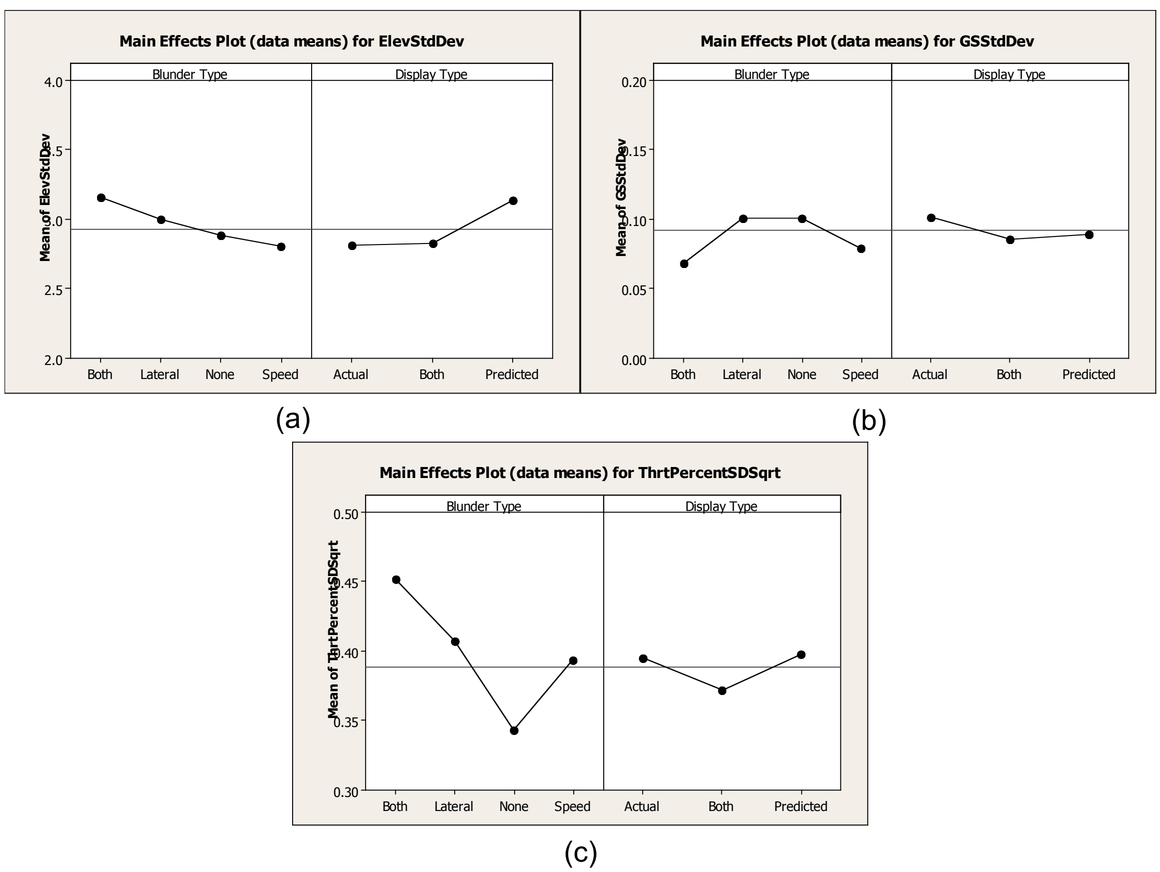


Figure 8. Least square means for control movements and glideslope deviation.

Regressions were run to determine if the subject pilots' were correlating their throttle movements with lead aircraft position, safe zone position, or lead aircraft speed. Regressions for the lead aircraft position and safe zone position were not significant. However, the regressions for lead aircraft speed show significant correlations, as shown in Table 2.

Table 2. Regression p-values for throttle movements by scenario and subject.

		Subject											
Scenario type	Overall	1	2	3	4	5	6	7	8	9	10	11	12
No Noncompliance	**0.000**	**0.004**	**0.014**	**0.020**	**0.000**	**0.001**	**0.000**	**0.005**	**0.040**	0.289	**0.000**	**0.000**	**0.025**
Pre Speed Noncompliance	**0.001**	0.890	**0.002**	**0.083**	0.148	**0.016**	0.112	0.917	0.119	0.524	**0.068**	**0.013**	0.403
Post Speed Noncompliance	**0.000**	0.485	**0.060**	0.526	**0.002**	**0.055**	**0.000**	**0.002**	**0.012**	0.331	**0.000**	**0.001**	**0.019**
Pre Lateral Noncompliance	**0.010**	**0.002**	**0.062**	**0.079**	**0.001**	**0.000**	0.438	**0.000**	**0.025**	0.544	**0.000**	**0.000**	0.239
Post Lateral Noncompliance	0.142	**0.086**	**0.012**	0.936	**0.037**	**0.000**	0.622	**0.000**	**0.025**	0.538	**0.028**	**0.001**	0.717

Discussion

There are several consequences of these findings. First, since the predicted safe zone is not accurate when the lead aircraft is violating the approach procedure (such as departing its approach course), a logical strategy for the pilots when given only the predicted safe zone display would be to monitor for noncompliance and initiate a go-around should it be detected. Conversely, since the actual safe zone does not depend upon the lead aircraft's compliance to procedure, a logical strategy would be to track the safe zone and not be concerned with the other aircraft's compliance. However, the lack of significant results for detection of blunders indicates that they were probably not using these strategies.

It appeared instead that pilots were monitoring for compliance regardless of display type. All twelve pilots mentioned and made frequent use of the text indication of speed on the navigation display, and four pilots indicated that a lateral trend indicator would be useful. These observations suggest that the pilots were concerned about compliance, but this concern was (incorrectly) independent of the display of the safe zone. It seems that the pilots were worried about the overall safety of the approach ("Am I closing on the lead aircraft? Is the lead aircraft turning towards me?") rather than about the effect of compliance on the safe zone.

More generally, the pilots' use of the displayed position of the safe zone appeared to be independent of the conceptual basis for the display. Eight pilots waited until after leaving the safe zone before executing a missed approach, despite it being inevitable in cases where their speed exceeded the lead's speed. Three other pilots began the missed approach procedure once it was inevitable that they would leave the safe zone, and the remaining pilot initiated missed approach once he perceived that the lead aircraft was not complying with the procedure in some way. In addition, when the lead aircraft departed its approach course, many pilots chose to remain within the (now inaccurate) predicted safe zone, just as they had done when presented with just the actual safe zone (which was still accurate under these circumstances). These findings suggest that the pilots were using the display of the safe zone as a "go/no-go" indicator; if they could remain within whichever safe zone was displayed they would continue on the approach, otherwise they would go around. This lack of distinction concerning the conceptual basis for the safe zone may have been encouraged by the similar format used for their presentation, which differed only in color (predicted was green, actual was yellow).

Another logical strategy when both safe zones were displayed would be to use the predicted safe zone to stabilize position within the safe zone. As essentially a "worst-case" safe zone, remaining within the predicted safe zone would also ensure that they remained within the actual safe zone for the duration of the approach. In fact, only two pilots made specific attempts to remain behind the predicted safe zone brackets when both brackets were available, and two other pilots indicated that they felt that the predicted safe zone display was not useful. This may be because the pilots felt that they could remain behind the actual safe zone without difficulty, although the actual safe zone could change length and position rapidly when the lead aircraft maneuvered.

There were a number of additional interesting observations concerning the participant pilots' compliance to approach procedures. First, there was a significant time lag between exiting the safe zone and initiating a go around, even in cases where the subjects could predict that they would depart the safe zone. In addition, one pilot chose to remain on his approach path in several scenarios despite leaving the safe zone, as he felt it unsafe to perform a go-around.

Two other pilots elected to not fly the missed approach as published when the other aircraft had crossed in front of them (and therefore was flying in the same direction as the missed approach course). Pilots seemed to be searching for additional confirming (or dissenting) information to support their decision. This observation is abundantly confirmed in a number of studies concerning compliance – pilots cannot be assumed to act automatically to procedures or alerts (Ockerman & Pritchett, 2000). However, the uncertainty of this environment may have justified the delays and the deviations from normal go-around procedures.

Another interesting result is that pilots have a wide variety of methods of maintaining spacing for current visual closely spaced parallel approaches. Each of the twelve pilots had a different target separation, ranging from nose-tail to 5 miles. Other responses included keeping the other aircraft at a 45-degree angle and keeping whatever spacing ATC had provided. Four pilots mentioned keeping vertical separation from the other aircraft. These strategies are problematic, as, for example, maintaining nose-tail separation is demonstrably unsafe - in this position a sudden turn by the lead aircraft can cause a loss of separation within ten seconds.

An obvious consequence of these results is that relying on compliance may not be advisable. A less obvious consequence is that pilots were expecting deviations from procedures. They therefore monitored information that related to particular deviations, but still made many errors in interpreting that information. Successful detection of noncompliance, moreover, did not translate into the pilots interpreting the consequences of that noncompliance.

The monitoring that pilots used for this approach appeared to be related to a strategy which may not have been ideal, but was sufficient and easily implementable. This strategy required only basic position, speed, and trend information for the other aircraft. Pilots generally complied with the instructions concerning remaining within the safe zone, but based their judgment of safety on their own strategy, despite understanding that the safe zone represented a better model of safety.

Conclusions

During training, pilots indicated an understanding of the different bases of the displays of the safe zone. In one case the validity of the safe zone indication was not dependent upon noncompliance and the indication could continue to be followed; in the other case the indication was not valid if the lead aircraft violated the procedure. In both cases, however, pilots chose to use a heuristic speed-matching strategy and did not react consistently or correctly to lead noncompliance. Pilots did not always comply with the procedure – in some cases their violations were ill-advised, in others their actions may have been safer than complying with the procedure. In either case, pilots may continue to use procedure information after the procedure no longer applies.

These results suggest that support for procedure following could assist pilots when flying approaches, and that this support should extend beyond the bounds of the nominal situations for which the procedure has been shown to apply. Such support should provide assistance in not only following procedures and understanding the context of the procedure, but also for how to proceed when the confines of the procedure have been departed.

Chapter 5

Using Procedure Context to Develop Information Displays to Support Operator Procedure Following

The results of the preliminary experiment suggested that a more rigorous examination of how to utilize procedure information was required. Pilots tended to follow a heuristic strategy, and, although concerned about noncompliance and trained on its effect, they were unable to translate this training into action. As discussed in Chapter 3, procedure context information may assist in providing support for this purpose.

In addition, the use of procedure context allows a shift away from the traditional focus on supporting compliance to the interactions called for in the procedure. It defines the larger role of the procedure within the system's environment, providing information to the user about how the specific actions in the procedure affect, and are affected by, the environment.

Procedure Context for Procedure Turn IAP

The guidance provided for procedure turns is discussed in Chapter 2, and pilots only have reference to their training and the approach plate while executing the procedure. However, there is considerable context about the procedure to which they do not have access (and for which they probably have not been trained). In addition, the complexity of the procedure may make some of the context of the procedure, for which a pilot has been trained, inaccessible. Following is a description of each of the elements of procedure context as it relates to procedure turn IAPs.

Intention

The intention of the procedure turn IAP is to transition the aircraft safely and consistently from enroute airspace to a point at which a descent to landing can be made.

Rationale

Each segment of the approach can be considered a step of the IAP. For each segment, it is important that the pilot understand the rationale for that part of the IAP. Table 3 below outlines the major segments of the procedure turn IAP, and the rationale for that step.

Table 3. Rationale procedure context for procedure turn IAP.

Segment	Rationale
Cross the IAF	Marks beginning of approach and identifies a place to begin turn outbound.
Turn outbound	The fly-off outbound enables the pilot to lose altitude and get set up for a consistent intercept of the final approach course.
Inbound turn altitude restriction	Keeps aircraft separated from terrain while in protected airspace but not on final approach.
FAF altitude restriction	Keep aircraft above terrain until established on final approach.
Timing	The MAP is identified as a distance from the FAF, but no distance measuring equipment is available; timing is therefore used to identify the MAP.
MDA	This is the minimum altitude on final approach that keeps the aircraft separated from terrain objects.
MAP	This is a point at which it has been determined that a safe landing can no longer be made from the MDA.

Boundary Conditions

Procedure turn IAPs are designed in accordance with FAA Order 8260.3B, titled "United States Standard for Terminal Procedures" (TERPs). This document mainly specifies the airspace restrictions within which the procedure must be confined (i.e. the protected airspace). These restrictions form a significant portion of the boundary conditions for procedure turn IAPs, but do not form a significant part of training, pilot reference documents, or displays. Figure 9 below indicates the protected airspace for a procedure turn.

Remaining clear of terrain obstacles is one of the main goals of IAPs, and this is guaranteed only when the pilot (1) remains within the airspace restrictions identified for the approach, and (2) remains above the specified altitudes for the particular portions of the approach. The latter is specifically identified on the approach plate and is (fairly) easily confirmed on the aircraft's instruments. The latter, however, is not clearly identified on the approach plate nor easily confirmed on the instruments.

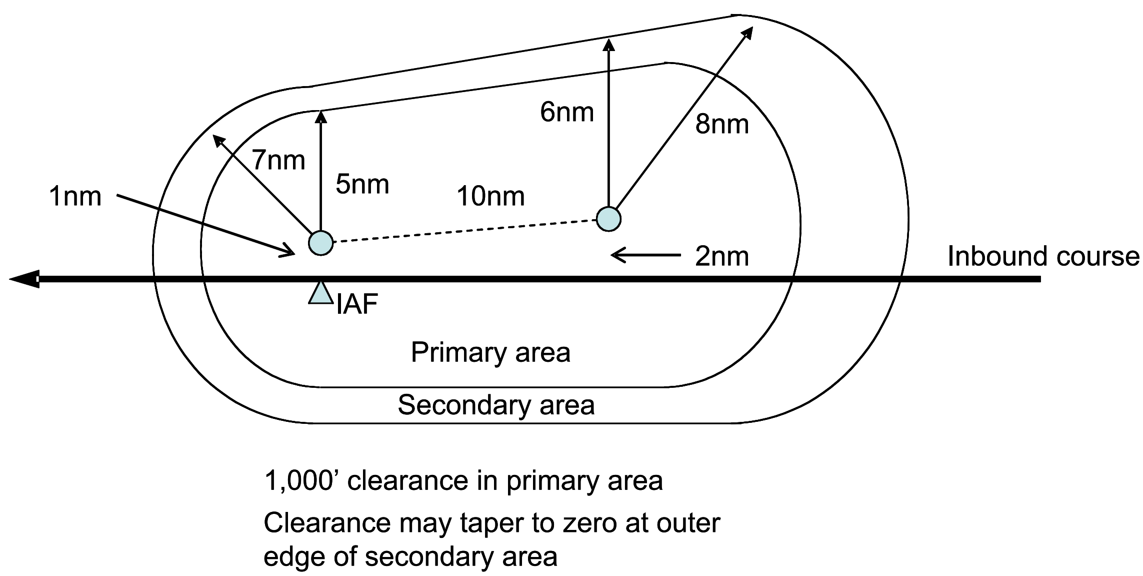


Figure 9. Protected airspace for IAP.

Triggering Conditions

Each step in Table 3 has a triggering condition. These are shown below in Table 4. The triggers indicate the events that must occur in order to transition from one step or segment to the next, or the criteria that, when met, releases the pilot from the restriction.

Table 4. Triggers for IAP segments.

Segment	Trigger
Cross the IAF	Station passage as indicated on navigation instruments
Turn outbound	After station passage
Inbound turn altitude restriction	Until established on the final approach course
FAF altitude restriction	Until passing FAF
Timing	Start at FAF
MDA	Maintain upon reaching until runway in sight or MAP
MAP	Upon expiration of timing

Temporal Construct

Due to the navigation instruments used for the procedure turn IAP shown in Figure 1, the pilot does not have access to any information regarding distance from the IAF or FAF. Since the aircraft is (in this case) instructed to remain within 10 miles of the IAF, and needs to know when the MAP is reached (which is only identified by it being 3.7 NM from the FAF), timing must be used in lieu of distance measurements. Pilots remain within 10 miles by flying away from the IAF for only a predetermined amount of time (generally 1 to 2 minutes). Pilots identify the MAP by using the timing block on the approach plate, which identifies different times for a number of groundspeeds.

The groundspeed for the approach differs depending on the aircraft type and pilot technique, and also will vary over the course of the approach due to imprecision and winds. As a result, the actual distance flown for a given time will vary.

For the fly-off from the IAF, it is only important to remain within 10 NM, but it is nearly impossible for the pilot to be sure of what distance has been flown without distance information. To determine this, the pilot would have to mentally integrate the distance flown given the variation in ground speed. Of course, an estimate could be made based on an average speed and time. For an aircraft flying 120 knots groundspeed (2 miles per minute), in 2 minutes the aircraft will go 4 miles. If this estimate is off by 20 knots, the distance estimate would be off by 2/3 of a mile. This is not significant unless the limit is to stay within 5 NM (which is not uncommon for approaches used by general aviation aircraft). For faster aircraft, 2 minutes at 4 miles per minute (240 knots) is 8 miles, and if off by 20% the aircraft would be within 2/5 NM of exceeding the 10 NM limit.

For identifying the MAP, pilots start timing when passing the FAF, and use either the time associated with the expected final approach speed. If the expected final approach speed is not one of the entries on the timing table, pilots typically use the closest time or interpolate (often crudely) to get a closer time. This estimate becomes less accurate the more the groundspeed varies as the pilot flies the final approach.

Without any automation support, however, it is impractical for pilots to continually adjust these times based on actual groundspeed, to try to more accurately estimate groundspeed, or to more accurately interpolate fly-off times.

Ordinality

Certain steps are required prior to initiation of others. For the procedure turn IAP, the sequence is generally obvious – it is either illogical or physically impossible to do some steps before others. For example, it is illogical to perform a course reversal after passing the FAF. However, it may also seem illogical to turn back outbound if the final approach course is intercepted, yet this is required once the IAF is passed. The reason for the outbound turn in this case is not to better align the aircraft with the final course but to enable the pilot to lose altitude and also to keep separation with other aircraft on the approach ahead of the pilot. The ordinality constraint here should be identified, and, if possible the reason for the constraint should be made clear.

An example of where ordinality is occasionally violated is on the inbound turn and its associated altitude restriction. The aircraft must be established on the inbound course prior to descending below that altitude (Federal Aviation Administration, 2002). Descending prior to establishing on course inbound may be dangerous in that full terrain separation assurance may not be provided. Yet it is not uncommon for pilots to descend sooner than is allowed by the procedure.

Necessity

For the procedure turn IAP, the only necessary items are the crossing of the IAF, the direction of turn, the altitude restrictions, the inbound course, the crossing of the FAF, and the identification of the MAP (Federal Aviation Administration, 2002). Other aspects of the approach are techniques for complying with these necessary items.

Reversibility

Errors during the IAP may or may not be reversible. Dropping below an altitude restriction can be reversed by climbing back above it, assuming terrain does not intervene. Course errors are (hopefully) always being reversed. If position errors become extreme, however, it may not be possible to reverse them and resume the approach.

If navigation errors become large, re-intercepting course in time to complete the approach may be too difficult to complete safely. In this case the pilot should abandon the approach. The guidance on how to abandon the approach is given by the missed approach procedure, however, and it typically relies on being on or near the approach course.

Appropriate Specificity

Much of the procedure turn IAP is non-specific; numerous methods are available to accomplish the procedure. This ambiguity is deliberate – the entry may be different depending on the approach to the IAF, and it is desirable to accommodate a range of techniques whose relative merits can be evaluated by the pilot under her or his specific circumstances.

The altitude restrictions depict their specificity. In the example provided (Figure 1 in Chapter 2), the restrictions are "at or above", which is indicated by the line below the altitude. If the restriction were meant to be "at" only, then there would also be a line above the altitude. Similarly, if the restriction were meant to be "at or below", then there would only be a line above the altitude. Altitudes that are simply recommended appear with no lines (Federal Aviation Administration, 2002).

The timing block appears specific when in fact it is not. Fluctuations in groundspeed and variations in where timing is started (relative to the FAF) will change the time it takes to fly to the MAP. However, without more specific guidance, the pilot must rely on the timing estimate.

Previous Step and Following Actions

The sequence of the IAP is fairly well defined, but many individual items need to be accomplished at nearly the same time. For instance, when departing the FAF, pilots need to begin timing, they may need to turn to a new final course, they need to establish a descent rate, and they may need to make a radio call to announce passing the FAF. Omitting any of these steps can lead to dangerous or unsafe situations.

Locational Indication

The procedure has two interconnected axes – the lateral profile of the procedure and the vertical profile. Pilots can, with difficulty, identify their position on the procedure for these two axes. The workload to do so, however, is typically higher than pilots can manage while also flying an instrument approach. Pilots flying the approach shown in Figure 1 (assuming they do not have access to a GPS display) have only an NDB to determine their position, and would be significantly taxed to identify their position with respect to the procedure airspace while manually controlling the aircraft.

The NDB approach is the IAP with the least informative navigation instruments. Most approaches utilize better navigation aids and instruments. In these cases pilots would have better locational procedure context, which may contribute to those approaches being less error-prone.

Forking

Forking in the procedure occurs when the aircraft must go around. At that point a transition to the missed approach procedure must be accomplished. Pilots must go around if any of the following are true:

- when reaching the MAP if the runway cannot be seen or a safe landing cannot be made,
- anytime it is determined a safe landing cannot be made, or
- instructed to do so by ATC (Federal Aviation Administration, 2002).

As mentioned previously, the MAP on the NDB approach is identified by timing, which makes its identification imprecise. Yet the missed approach procedure is designed with the assumption that the aircraft is at the MAP. Beginning the missed approach procedure away from the MAP can impact the ability of the aircraft to complete the procedure safely or cause separation problems with other aircraft.

Displays for Procedure Turn IAP

Having identified the procedure context for the procedure turn IAP, one can then identify useful content for displays to support procedure following. Of interest to this research is identifying context information whose current presentation is inadequate. Testing whether displaying that information enhances pilot performance will then provide evidence of its utility. Table 5 below shows the context elements, and how those elements are currently made available to the pilot.

Table 5. Procedure context elements in the procedure turn IAP.

Element	***Where found in procedure turn IAP***
Intention	Training
Rationale	Training, FARs, TERPs, Approach plate
Boundary conditions	TERPs
Triggering conditions	Approach plate, instruments, visual contact with runway
Temporal construct	Clock, timing block, groundspeed
Ordinality	FARs, Training, Approach plate
Necessity	FARs, Approach plate
Reversibility	Instruments
Specificity	Approach plate, FAR
Previous/Following actions	Training, FAR, Approach plate
Location indication	Instruments, Approach plate
Forking	Instruments, Approach plate

Of the elements in Table 5, several could benefit from being represented in some way to the pilot other than the manner in which it is currently done. The rationale for remaining within 10NM, for the turn to be conducted on a particular side, and for altitude restrictions is given by TERPs criteria. However, only minimal elements of this rationale are provided to the pilot on the approach plate. A more specific depiction of this rationale, which could be accomplished by explicitly identifying the protected airspace, would be extremely useful for the pilot to see when selecting maneuvers or when the procedure can no longer be followed in the standard manner.

The temporal constraints on identifying the MAP are given by the appropriate groundspeed entry on the timing block on the approach plate, and referenced by a clock. This is not ideal because the actual position of the MAP is based on distance and not time. In order for the pilot to convert to time, a particular groundspeed must be chosen (although actual groundspeed is dynamic). The pilot must then start the clock at the correct location, and identify which time matches the entry for that groundspeed in the timing block. If the approach groundspeed falls between entries, interpolation may be required. For identifying the MAP, this series of estimates and somewhat mentally taxing operations could be replaced with a more specific, more informative, and more accurate countdown to the FAF. One simple improvement would be to perform the interpolation for the pilot, and display the time left to the MAP.

Altitude restrictions appear on the approach plate, but much of the context (necessity of the altitude restriction, triggering conditions for passing an altitude restriction) is not clearly provided to the pilot. Visual triggers of different vertical phases of flight and confirmation of the necessity of those restrictions could be identified by highlighting the currently appropriate altitude restriction.

Other than through a difficult mental transformation of information on the navigation instruments, the pilot's location within the procedure is not clearly specified anywhere on the flight deck. A location indication of the aircraft's position within the procedure should therefore be useful to the pilot.

Since much of the above information is dynamic, it cannot be solely addressed through training or a static presentation of the information (such as on an approach plate). The information, then, may either be presented on an existing dynamic display, or a new display could be developed. For experimental purposes, it is desirable to be able to isolate the effects of adding the context information; minimizing the use of new displays and new symbology is therefore preferred.

There are three distinct items described above: timing, altitude restrictions, and procedure location. The first two may be simply displayed as text or graphically on existing instruments. The last requires a display similar to the approach plate itself, which would require the addition of a new map-like display.

What must be determined, then, is how to display timing information, altitude information, horizontal location information, and procedure location. For an operational display, human factors and HCI principles would be applied to design these displays. For purposes of this study, the specific design considerations for an operational display are left to future studies. The distinction that is to be tested is whether the addition of displayed elements of procedure context is useful to the operator; as such, a display that effectively isolates these elements is needed. The resulting display, described in Chapter 6, may be not be ideal from an HCI perspective, but needs to provide visibility as to whether the operator is benefiting from the addition of these display elements as opposed to anything else that is being presented.

Chapter 6

Dissertation research - Method

In a preliminary experiment described in Chapter 4, commercial airline pilots flew a new procedure for executing an instrument approach with another aircraft flying a simultaneous approach to a closely spaced parallel runway. That study indicated that pilots expect noncompliance, and actively search for information to confirm compliance. However, pilots were unable to interpret the consequences of noncompliance. In addition, pilots would sometimes actively and intentionally depart from the procedure if they felt that it was unsafe. That experiment suggested that a more rigorous examination of how procedure information could be used to support procedure following was needed.

It is suggested here that procedure context elements can provide guidance for how to provide such support. The previous chapter identified procedure context elements that, if displayed, might benefit pilots flying procedure turn IAPs. These elements were incorporated into a display for purposes of evaluating their effect on procedure-following.

In the dissertation research discussed in this and the following chapters, participant pilots flew an instrument approach procedure (IAP) with an electronic display of the procedure. In some cases this display contained only the information currently provided on the approach plate, while in other cases the pilot was provided with additional procedure context. Pilot performance, workload, and situation awareness while flying the approach procedure was measured, and statistical tests were applied to determine if the addition of the displays changed these measures. This chapter will detail the task that the participants were asked to perform, the equipment used for the experiment, the participants, and measures.

Experimental considerations

The purpose of this research is to test the belief that procedure context elements support procedure-following; it is not the purpose of this research to develop an operational display. As such, experimental considerations will take precedence over an implementation that would be operationally viable.

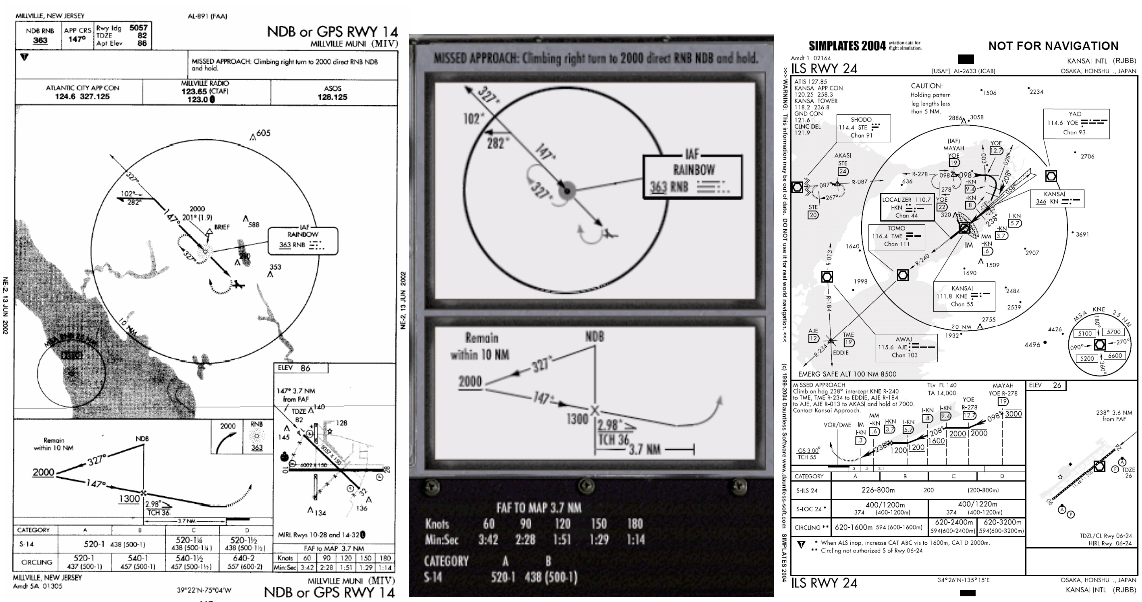
The dissertation research tested the elements discussed in Chapter 5. These elements were such that they could be incorporated into the display without providing benefit to the pilot in the form of improved navigational information or significantly reduced workload. For an operational system, these limitations may be a requirement; nonetheless the analysis would recommend what elements would best provide support, and what equipment and form such elements should take.

Experiment design

Participant pilots in this experiment flew a procedure turn IAP to Millville Municipal (an airport in New Jersey). This approach and airport were chosen due to the availability of a procedure turn approach, its use of an NDB without distance information, the low likelihood that participant pilots (who were located in California) would have flown the approach, and that the airport can be considered to be uncontrolled. These factors added the desired level of difficulty and realism.

Pilots flew this procedure using a baseline display (electronic format approach plate without procedure context elements, similar to that shown in Figure 1) in half of the trials, and flew with an enhanced display (the same electronic format approach plate with procedure context elements added) in the other half.

The baseline display was designed to display the same information as the paper copy with which the subjects were familiar, and the enhanced information was an overlay of additional information (it did not remove any of the standard information). During training the display was discussed in detail to ensure that the pilots knew where information could be found on the display, and did not have any questions. Interestingly, a few subjects noted that the display was based on the FAA version of printed approach plates rather than those sold by Jeppesen Sanderson, Inc (called Jeppesens). A comparison of an FAA version, and the Jeppesen version of the same approach is shown in Figure 10. As can be seen from the figure, the two versions are very similar, and after a review of the electronic version, no subjects noted any difficulty in interpreting it.



FAA/DOD version (a)

Experiment version (b)

Jeppesen version (c)

Figure 10. Comparison of experiment approach plate with Jeppesen and FAA versions.

The enhanced display did not assume any information other than that available to the other instruments in the aircraft (e.g. it did not use absolute position information such as provided by GPS). This was to ensure that any benefit provided by the enhanced elements was not due to the addition of information about the environment.

A seventh, deviant scenario was added for subjects 7 through 16. This deviant scenario required the subjects to unexpectedly have to fly the approach engine-out, requiring them to intentionally not comply with the approach procedure. This was desirable since one of the goals of the experiment was to investigate intentional noncompliance, and the expected intentional noncompliance was not being seen in the trials with the first six participants.

To introduce complexity, the entry angle for the procedure turn was varied. The entry angles used were designed to provide the participant with clear guidance as to the turn required, ambiguous guidance, or conflicting guidance. To examine whether the display aid was useful for the missed approach portion of the approach, the weather was lowered such that the runway would not be seen at the minimum descent altitude, forcing the pilot to fly a missed approach.

Hypotheses and variables defined

The hypotheses of the experiment are as follows:

> H_{011}: The presence of procedure context elements on the display does not affect situation awareness during nominal operation.
>
> H_{A11}: The presence of procedure context elements on the display affects situation awareness during nominal operation.
>
> H_{012}: The presence of procedure context elements on the display does not affect situation awareness during off-nominal operation.

H_{A12}: The presence of procedure context elements on the display affects situation awareness during off-nominal operation.

H_{021}: The presence of procedure context elements on the display does not affect workload during nominal operation.

H_{A21}: The presence of procedure context elements on the display affects workload during nominal operation.

H_{022}: The presence of procedure context elements on the display does not affect workload during off-nominal operation.

H_{A22}: The presence of procedure context elements on the display affects workload during off-nominal operation.

H_{031}: The presence of procedure context elements on the display does not affect safety during nominal operation.

H_{A31}: The presence of procedure context elements on the display affects safety during nominal operation.

H_{032}: The presence of procedure context elements on the display does not affect safety during off-nominal operation.

H_{A32}: The presence of procedure context elements on the display affects safety during off-nominal operation.

The measures available to test these hypotheses are shown in Table 6.

Table 6. Hypotheses and measures.

Hypotheses	*Measure*	*Units*
H_{011}- H_{A12}	Errors in subjects' recollection of ground track	Number of errors
H_{011}- H_{A12}	Errors in subjects' recollection of speed deviations	Number of errors
H_{011}- H_{A12}	Error in subjects' recollection of altitude deviations	Number of errors
H_{011}- H_{A12}	Errors in subjects' recollection of missed approach deviations	Number of errors
H_{021}- H_{A22}	NASA TLX workload ratings	1-24 scale subjective rating
H_{021}- H_{A22} H_{031}- H_{A32}	Variance of control inputs	Degrees of bank and pitch
H_{031}- H_{A32}	Altitude errors	Number and duration of errors
H_{031}- H_{A32}	Track errors	Number and duration of errors

The variables and levels in the experiment are:

A. Subjects (blocking variable)

B. Presence of procedure context elements on the display

    i. Present

    ii. Not Present

C. Entry into procedure turn

    i. Clear

    ii. Ambiguous

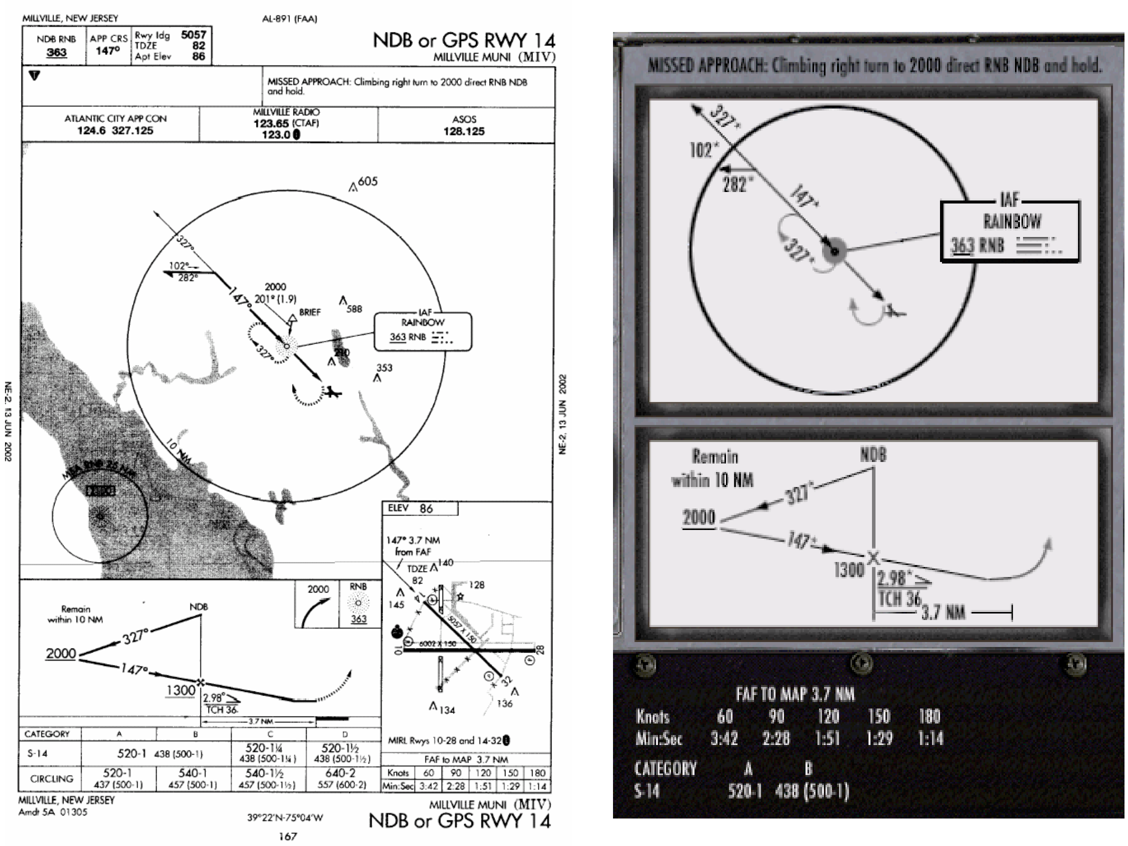
    iii. Difficult

The experiment represents a randomized 2-factor block design, with one variable ("display") having two levels, and the second variable ("entry") having three levels. As mentioned previously, the last nine subjects flew an additional deviant scenario, which will be considered separately from the other scenarios. Each pilot flew two or three practice approaches and then six (seven for those subjects who flew the deviant scenario) trials representing each case of the 2x3 test matrix shown in Table 7 and described above (variables B and C). To counterbalance the effect of having the enhanced displays, half of the pilots flew the A trials first, and half of the pilots flew the B trials first. The order of the trials (within the A and B trials) was randomized.

Table 7. Experiment matrix

Condition	*Display*	*Entry*
A1	Baseline	Clear
A2	Baseline	Ambiguous
A3	Baseline	Conflicting
B1	Enhanced	Clear
B2	Enhanced	Ambiguous
B3	Enhanced	Conflicting
Deviant (subjects 7 through 16 only)	Random	Random

Pilots were provided with an electronic version of the approach plate for the NDB procedure turn approach to Millville Municipal Airport. The actual approach plate is shown in Figure 11a, the electronic version is shown in Figure 11b. Since the pilots were required to be instrument trained, each will have already been trained on how to fly a generic procedure turn IAP. A full description of how to fly a procedure turn is provided in Chapter 2. In general, since the aircraft may approach the procedure turn fix from any direction, the actual direction and extent of turns required vary by individual circumstance. The pilot must decide, based on the direction of entry, how to enter the protected airspace so that she or he accomplishes the course reversal and aligns herself or himself with final.



(a) (b)

Figure 11. Millville NDB Runway 14 approach plate and electronic version.

For this task, three initial positions were used, as shown in Figure 12a through 12c. One position (Figure 12a) was such that the aircraft approached the procedure turn fix conveniently aligned to perform a 45º-180º or a 80º-260º entry, so that a left turn to the outbound course began the procedure. From a second position (Figure 12b) pilots were conveniently aligned to perform either a teardrop entry or a holding entry, and could have either turned to parallel the course outbound, or turned 45º to the right to enter the protected airspace. From a third position (Figure 12c) no convenient entry maneuver existed, so that the pilot was unable to begin the procedure turn on the protected side unless some deviation from either ATC instructions or the instrument approach procedure was performed.

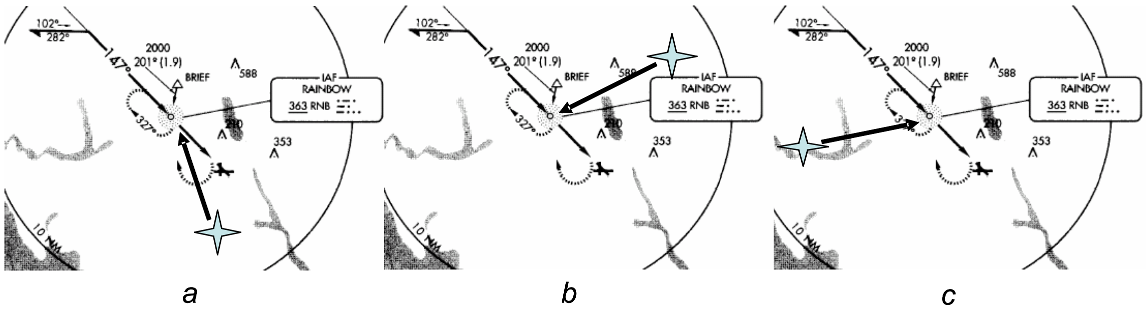


Figure 12. Starting positions and angles of entry.

Also included on the approach plate were missed approach instructions. These procedures are designed to transition the aircraft from the approach back into the air traffic flow so it can be resequenced. A close-up of the missed approach instructions for this task is shown in Figure 13.

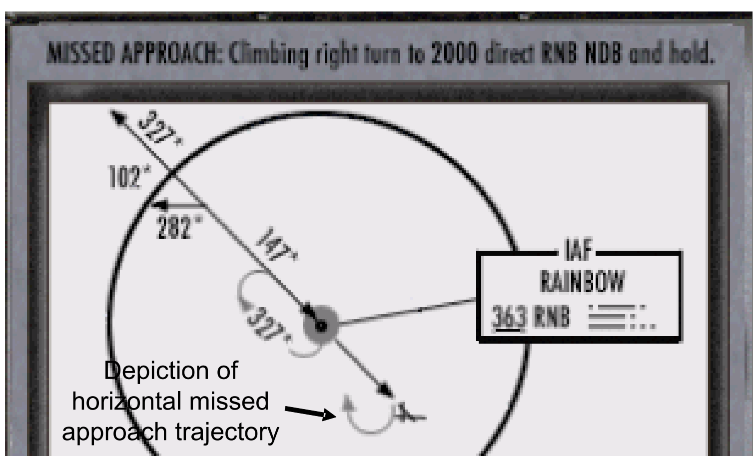


Figure 13. Missed approach instructions at top of approach plate.

The missed approach occurred at the missed approach point (Figure 14), as indicated by the expiration of timing which estimates the passage of 3.7 nm from the FAF.

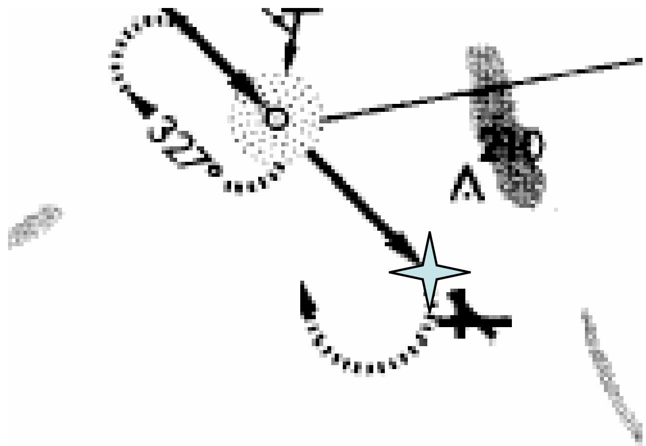


Figure 14. Location of missed approach

For the B set of trials, participants were provided with an enhanced display for the procedure turn IAP. This display contained the elements discussed in Chapter 5 and is shown in Figure 15.

Each participant was briefed on the equipment and the task, filled out a questionnaire regarding demographics, performed two practice approaches (3 subjects requested and flew one additional practice approach), and then performed 6 or 7 approaches corresponding to the experiment matrix. During each approach, data were recorded. The trials were videotaped as a backup to the data collection process. After each approach a questionnaire regarding the entry and missed approach was administered to test the participant's situation awareness, and then a workload questionnaire was administered.

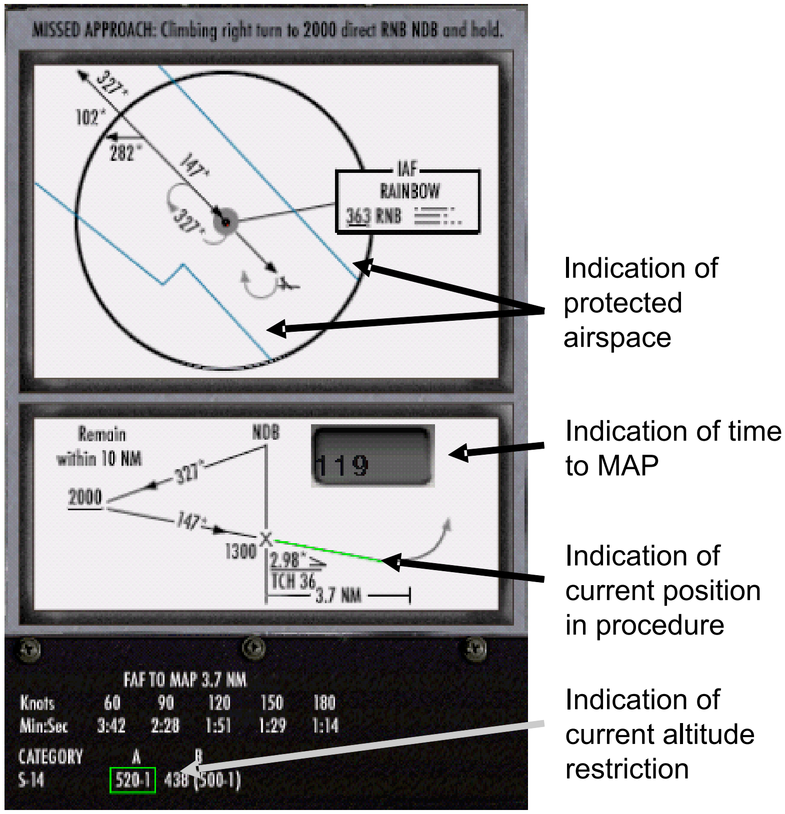


Figure 15. Enhanced display.

Equipment

The equipment setup is shown in Figure 16. The participant pilots flew a desktop flight simulator running on a 1.2 GHz IBM Thinkpad R31 laptop computer running Windows XP. The software used was Microsoft Flight Simulator 2002, with the standard flight instruments (except for the addition of the electronic approach plate) for the type of airplane used in the simulation. The simulation airplane was a Cessna 182 (instruments shown in Figure 17) was used for all trials. Procedure context elements were added to the electronic approach plate for the B trials.

Figure 16. Experiment setup.

Figure 17. Instrument panel Cessna 182.

A CH Products Flight Sim control yoke and throttle quadrant, and CH Products rudder pedals was used to control the aircraft. This control yoke includes controls for flap position, pitch and aileron trim, mixture control, and prop speed.

The simulator was set up for IFR flight to Millville Municipal Airport in New Jersey, with weather set at calm winds, overcast at 400 feet above ground level, and 10 miles visibility. This weather prevented the pilot from seeing any terrain features, and forced her or him to rely solely on instruments to perform the procedure turn and approach. The aircraft positions were preset to the positions described earlier, at an altitude of about 2,600 feet, and about 120 knots.

For subjects 7 – 16 the simulator was preset with an engine failure. Shortly before the aircraft passed the IAF the engine quit, requiring the pilot to try to turn towards the runway and land.

Participants

Participants were 16 private pilots from the West Valley Flying Club in Palo Alto, California who were recruited through advertisements at the club and on their newsletter, and also through email. Subjects were paid $50 for their participation, and no incentives for performance were offered. The advertisements stated that the pilots had to be instrument trained and comfortable flying procedure turn instrument approaches.

The experiment was conducted in a flight debriefing area at the club facility on the grounds of the Palo Alto Municipal Airport. Subject 13 was dismissed and his data was not used after he indicated he was not instrument trained. The remaining 15 subjects were instrument trained and indicated that they were familiar with the procedure turn IAP. Each volunteer was questioned regarding their level of knowledge of instrument procedures in general and of the procedure turn in particular. 3 of the 15 subjects were female, and 5 were instructors. One subject was air transport pilot rated, 4 were commercial, multi-engine rated, 2 were private, multi-engine rated, and the rest of the subjects (8) were private pilot rated. All but one subject were instrument rated; the one subject who was not was an advanced student recommended by his instructor (another subject). Age and hours, both total and instrument, were recorded, and are shown in Table 8.

Table 8. Age, total hours, and instrument hours for participants.

	Age	Total Hours	Instrument Hours
Median	35	700	90
Minimum	26	152	30
Maximum	51	13700	3000

Measures

Aircraft state data were recorded by Microsoft Flight Simulator then imported into Microsoft Excel, Minitab, and S-Plus. These packages were used to analyze the data. The state data recorded were aircraft position, speed, heading, altitude, and control actions. The names and descriptions of the data recorded are shown in Table 9.

After each trial, a questionnaire designed to test the pilots' situation awareness of what transpired was administered. This questionnaire is shown in Appendix 1. The subjects were asked to indicate on a map their initial position and ground track for the flight they just completed. Next, the pilots were asked to indicate their recollection of any deviations from the altitude requirements, the speed requirements, or the missed approach requirements of the procedure, and the reasons for those deviations.

Table 9. Data recorded

Name	*Description*
ADF_NEEDLE	Relative bearing of NDB
AIRSPEED	True airspeed in knots
ATTITUDE_INDICATOR_BANK_DEGREES	Indicated degrees of bank
ATTITUDE_INDICATOR_PITCH_DEGREES	Indicated degrees of pitch
ELAPSED_SECONDS	Number of seconds since start of simulation
ELEVATOR_TRIM	Degrees of pitch trim
ENGINE1_THROTTLE_LEVER_POS	Position of throttle
FLAPS_HANDLE_POS	Position of flap handle in degrees
ELEVATOR_POS	Position of elevator in degrees
RUDDER_PEDAL_POS	Position of rudder in degrees
YOKE_POS_X	Position of yoke for roll axis
YOKE_POS_Y	Position of yoke for pitch axis
PLANE_ALTITUDE	Altitude of aircraft in feet
PLANE_HEADING_DEGREES_GYRO	Indicated heading of the aircraft
PLANE_LATITUDE	Current latitude of aircraft
PLANE_LONGITUDE	Current longitude of aircraft
VERTICAL_SPEED	Vertical speed of the aircraft in feet per second

After the situation awareness questionnaires were completed, pilots were asked to fill out a NASA TLX workload questionnaire, also shown in Appendix 1.

The data described in Table 9 were used to produce several composite measures. The actual ground track of the aircraft was provided by the `PLANE_LATITUDE` and `PLANE_LONGITUDE` variables. Subjects' estimates of position as drawn on the map were compared with each subject's recollection of the aircraft's ground track for the following:

- starting position
- position crossing the IAF
- outbound track
- intercept
- inbound track

- position crossing the IAF
- final approach track
- position when starting the missed approach

The subject's recollection was scored as to being correct (1) or incorrect (0), and the sum of these 8 represented the subject's score for recall of lateral position.

The actual track of the flight was compared with the minimums for the approach to obtain a measure of flight error. Laterally the procedure required the pilot to remain on a bearing of 327º from the NDB when inbound to the FAF, then follow the 147º bearing out of the NDB until the MAP. Vertically the pilots were required to stay above 2000 feet until on the 327º bearing from the NDB inbound, stay above 1300 feet until the FAF, then stay above 520 feet for the remainder of the approach. The sampled position and altitude were compared with these limits for the appropriate parts of the approach to obtain a measure of error. Both the magnitude and duration of the error was calculated.

In addition, when flight error was detected, the pilot's correction back to the procedure was calculated. This was intended to provide an indication of when the error was recognized by the pilot.

The number of control inputs was extracted from the `GENERAL_ENG_THROTTLE_LEVER_1_POSITION`, `YOKE_Y_POSITION`, `YOKE_X_POSITION`, and `RUDDER_PEDAL_POSITION` variables. The values of these variables were plotted, and summary statistics determined. More frequent and greater extent of control inputs are an indicator of increased workload, and also provide a measure of the stability of the approach. Greater stability on the approach is desirable, and low stability is considered unsafe.

Chapter 7

Results

The measures described in the previous chapter were analyzed using Minitab and S-plus, with the following results. The deviant scenario will be described separately. For the analysis described below, results are found to be significant if the P-value of a test is below 0.05, and are found to be marginally significant if the P-value is between 0.05 and 0.10.

General results and observations

There were obvious differences between the participants' abilities, but all were able to fly the approaches without exiting the protected airspace for the approach. Every pilot successfully performed a reversal maneuver, flew inbound, descended to the MDA, and executed a missed approach. Additionally, each pilot attempted to maintain the proper course and altitude, and either corrected or mentioned deviations.

With one exception, all errors over 1.5NM on the non-maneuvering side of the approach airspace occurred with the baseline display. The exception was a case in which the subject was intentionally deviating from procedure; the subject had the same error while using both the baseline and enhanced displays. There were a total of nine such errors involving 5 subjects, and 7 occurred with the conflicting entry case. The largest error took the participant's aircraft 3.6 NM off course, 0.4 NM from the boundary of the primary obstacle clearance area.

The majority of the pilots flew the approaches in the same way, regardless of entry. Table 10 indicates that in 84 out of 90 cases, the 45-180 maneuver was selected. This was somewhat surprising, and resulted in less variation in the entry maneuver than was anticipated. It was expected that for the ambiguous case, many subjects would use the racetrack entry (only 1 subject did so, once with each display). It was also expected that for the conflicting case, numerous subjects would see the need to depart from the procedure (only 1 subject did so, once with each display).

Table 10. Maneuvers used by entry.

	Ambiguous	Clear	Conflicting	All
45-180	28	29	27	84
Adhoc	0	1	3	4
Racetrack	2	0	0	2

Ad-hoc maneuvers were maneuvers that did not fit into any of the normal techniques described in Chapter 2. Two of these cases were attributed to one subject (once with the baseline display, once with the enhanced display) who performed a racetrack on the wrong side of the inbound approach course for the conflicting entry scenario. This subject recognized that he was on the non-maneuvering side, continued outbound on that side, then reversed course (correctly, but in the opposite direction indicated by the procedure) back towards the maneuvering airspace. These wrong-way racetrack maneuvers represented an intentional violation of procedure. The other two ad-hoc entry maneuvers were essentially a hybrid of the 45-180 and 80-260 maneuvers which did not technically violate the procedure.

Despite the NDB being a very imprecise navigation aid, in 42% of the cases participant pilots were right on course at the point where they would have had to visually identify the runway. Only in 20% of the cases were the subjects clearly not in a position to land at the end of the approach, and the pilots were still heading roughly toward the runway in even those cases. Kruskal-Wallis nonparametric tests across the variables indicated a significant difference by subject (p=0.003), but not for display, entry, or order of runs.

Statistics of error between recollection of ground track and actual ground track

For this measure, a comparison is being made between the actual ground track and the recalled ground track. A score was compiled for each experimental run. A minimum of 0 was possible, and a maximum of 8 was possible. No subject scored below 2, and no question was always answered right or wrong. The data will be checked to see if significant differences exist by subjects, display type, or entry. A summary of the data is shown in Figure 18.

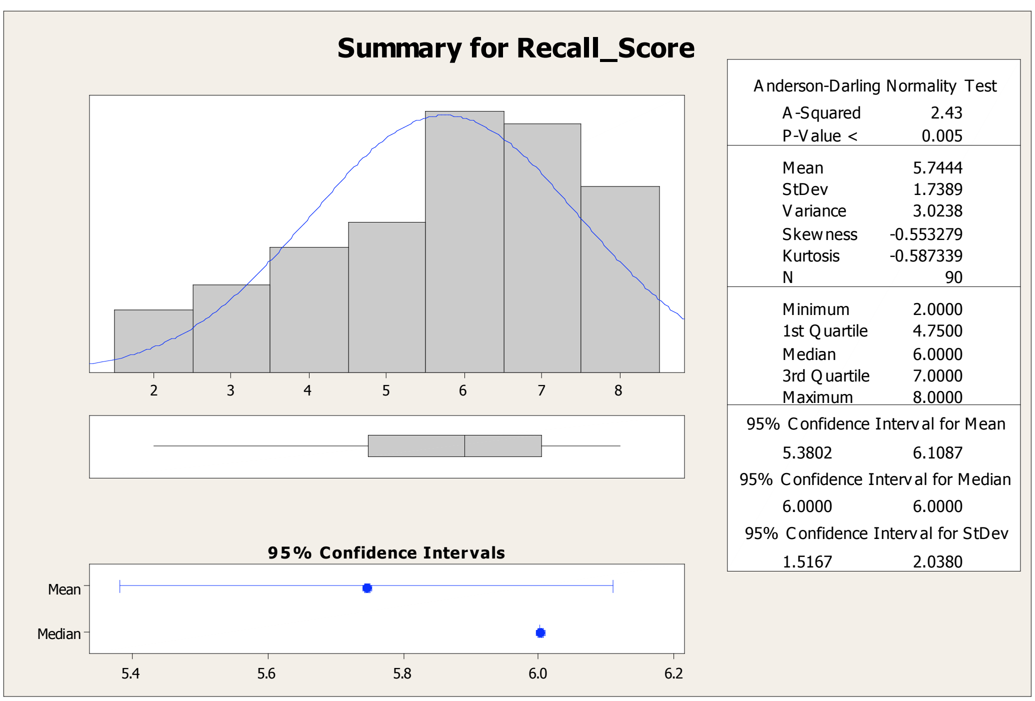


Figure 18. Summary of lateral_recall_score

The recall score is mildly right skewed (-0.55), and fails the Anderson-Darling normality test. No other serious anomaly is apparent for the data. A tabulation of the data is shown in Appendix 2.

None of the scores or groupings stand out significantly from the overall mean and median, and none have excessive or low standard deviations. A boxplot of the values for the subjects and independent variables is shown in Figure 19(a) through (c). Figure 19(d) shows the boxplot for order of runs to check for learning effects.

Figure 19(a) indicates some differences between subjects; particularly noticeable are the high scores of subject 10. Figure 19(b) does not seem to indicate any effect by the entry variable. Figure 19(c) shows a possible increase in scores for the enhanced display over the baseline display. Figure 19(d) does not seem to indicate any effect due to the order of runs. A few outliers are noticed on the boxplots, but these scores were checked for accuracy and were found to be correct, so they were not discarded.

Although the data did not pass the Anderson-Darlington normality test, both the Ryan-Joiner and Kolmogorov-Smirnov tests indicated normality ($p>0.1$), so an ANOVA was run. The departure from normality exhibited by the variable seemed mild, with a slight right skew and no large differences in variance, except perhaps for the scores of subjects 10 and 12, who had small variances in scores. Since the effect of the display variable may not be uniform across entry types, these variables were checked for interaction effects. Figure 20 shows that the display variable may not have had the same effect for each level of the entry variable, so the model did not include this interaction term. The results of the ANOVA are shown in Table 11.

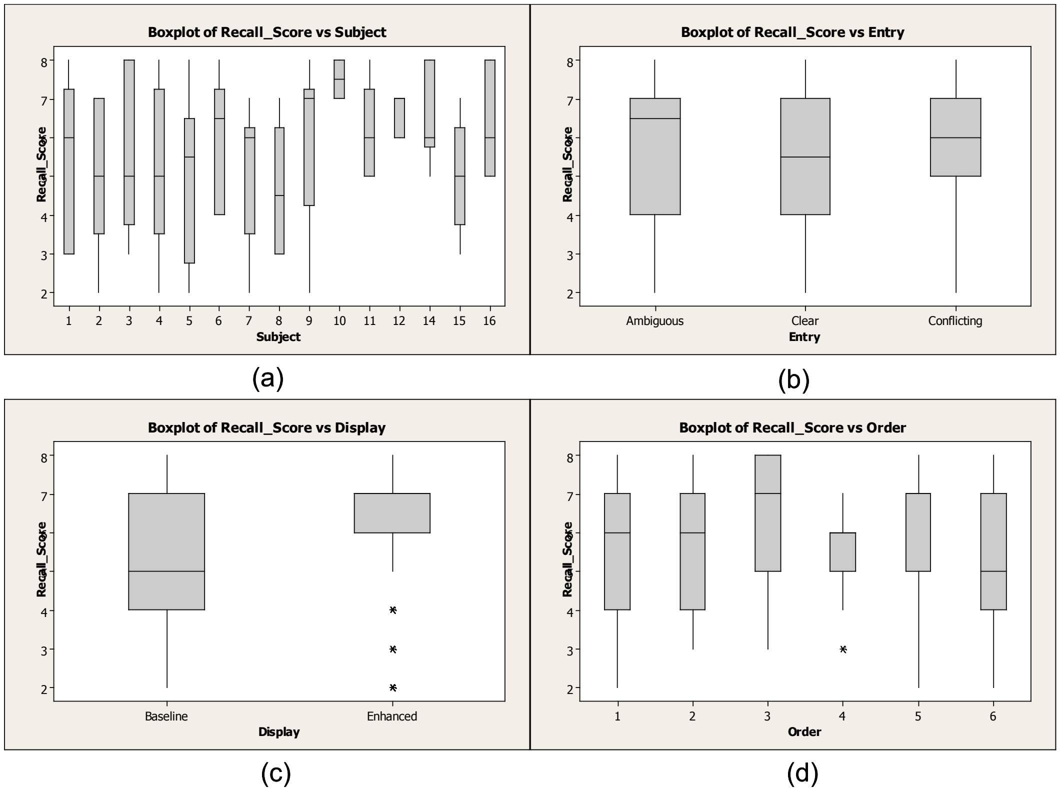


Figure 19. Boxplots of lateral score.

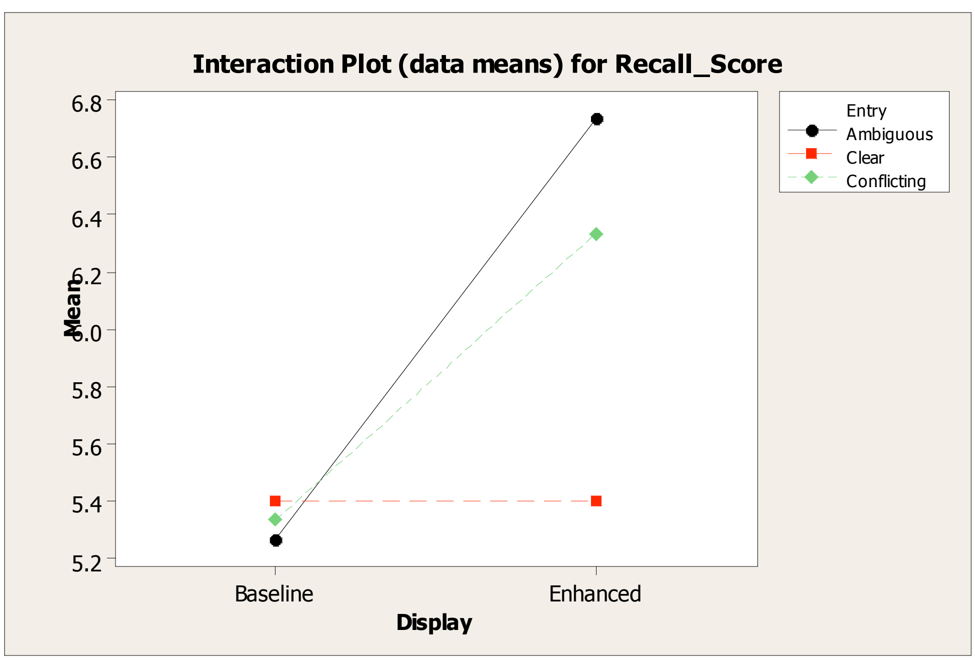


Figure 20. Interaction plot for display and entry variables.

The ANOVA indicates a significant difference across display (p=0.020), but not for subject, entry, or the interaction term.

Table 11. ANOVA for lateral recall score.

Analysis of Variance for Recall_Score, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	52.622	52.622	3.759	1.41	0.174
Display	1	15.211	15.211	15.211	5.69	0.020
Entry	2	5.756	5.756	2.878	1.08	0.346
Display*Entry	2	8.422	8.422	4.211	1.58	0.214
Error	70	187.111	187.111	2.673		
Total	89	269.122				

Figure 21 indicates that the mean recall score was higher for the enhanced level. Interestingly, the interactions plot (Figure 20) shows that most of this difference came from the improvement provided for the ambiguous and conflicting entry scenarios, although the results of the ANOVA could not reject the hypothesis that the interaction term was not significant.

A check of the residuals is shown in Figure 22. One point seems to be troublesome, with a very low (negative) residual. There is some indication of a reverse funnel shape to the residuals versus fitted values, which may be due to there being few high fitted values. The residuals seem to show no dependence on order (the data were ordered by the sequence of runs). These results suggest that caution should be used in interpreting the ANOVA. However, the one significant result is fairly strong, with _ < 0.025.

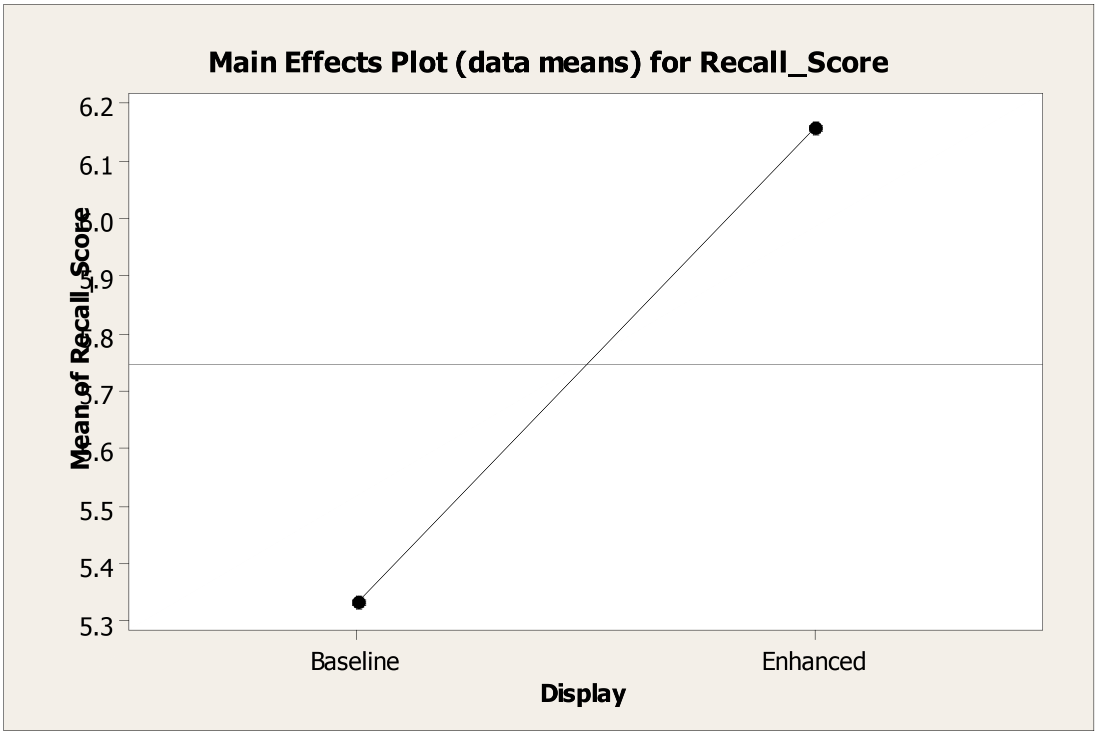


Figure 21. Main effects of display on lateral recall score.

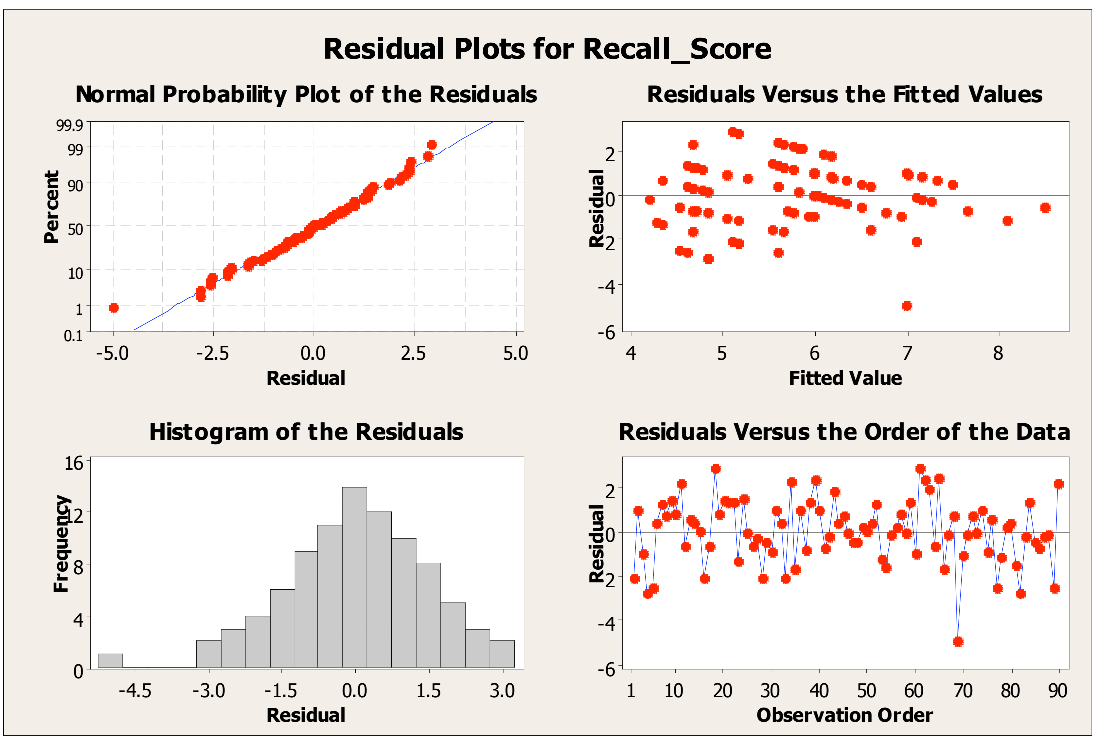


Figure 22. Residuals for lateral recall score ANOVA.

Error in recall of altitude deviations

The next variable to be checked is the error between the subjects' recall of altitude deviations. The subjects answered 3 questions about their compliance to altitude, which was checked against the record of their actual altitude profile. A minimum of 0 was possible, and a maximum of 3 was possible. No subject scored below 1, and no question was always answered right or wrong. The data will be checked to see if significant differences exist by subjects, order of runs, display type, or entry.

Figure 23 shows that the data is skewed right, and again does not pass the Anderson-Darlington test for normality. However, the data passes both the Kolmogorov-Smirnov and Ryan-Joiner tests for normality ($p>0.1$). Tabulated statistics are shown in Appendix 2.

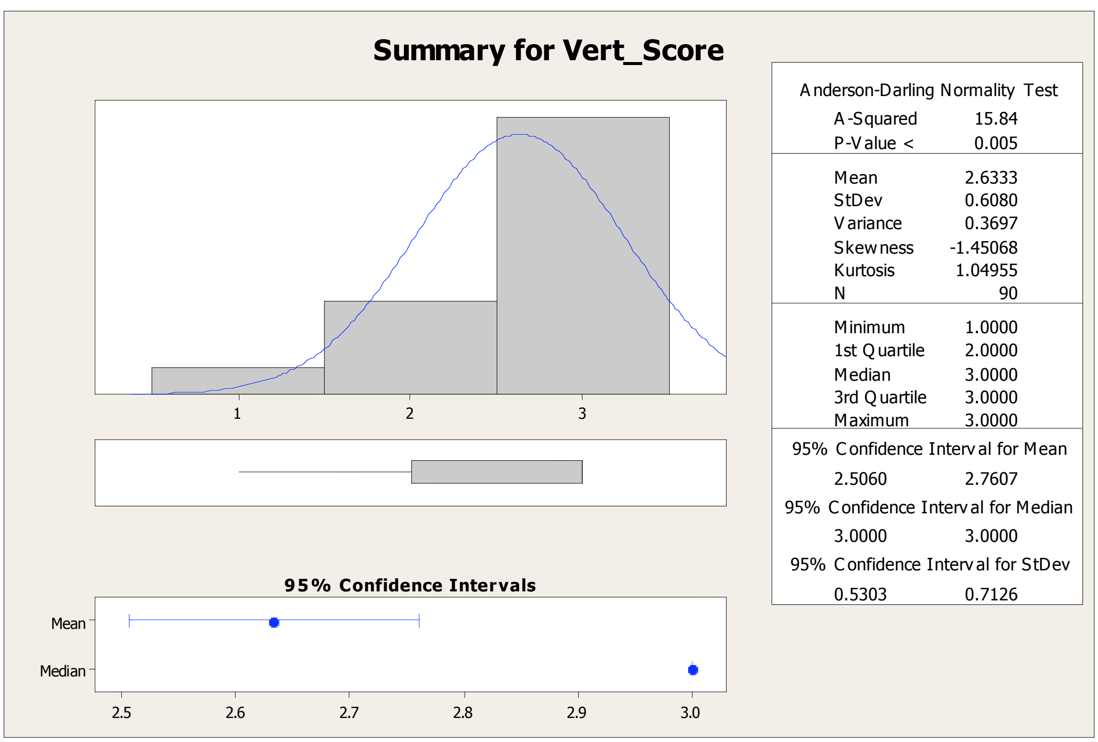


Figure 23. Graphical summary of altitude recall statistics.

The mean of all scores is between 2.4 and 2.933, although there are some differences in standard deviations (minimum is 0.2582, maximum is 0.8281).

This is itself somewhat interesting, suggesting that the variance for the enhanced display is lower and less variable than for the baseline display. This is probably caused by the maximum score being anchored at 3. If the performance using the enhanced display were better, then more scores would be at 3, fewer at 1 (or 0), and the variance would shrink.

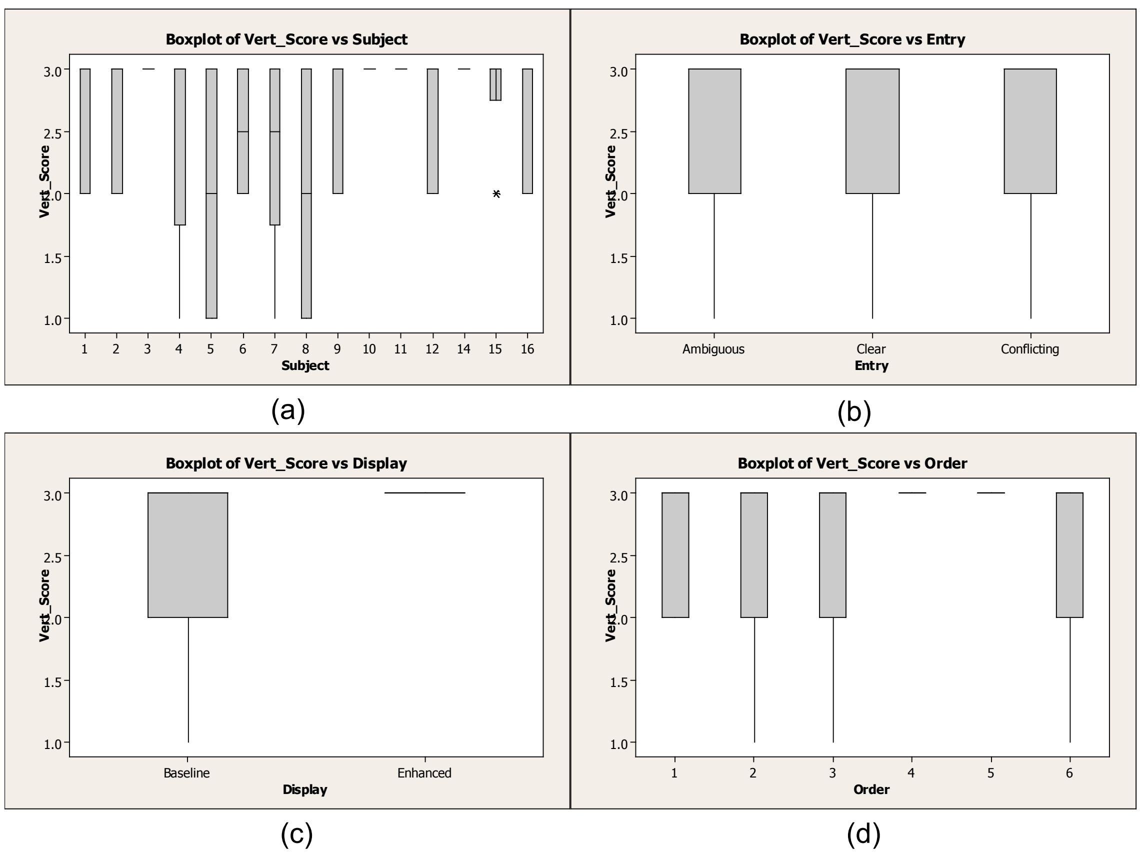


Figure 24. Boxplot of other recall scores by display.

Figure 24 seems to indicate some differences across subject and perhaps order, as well as for display. The interactions plot in Figure 25 seems to indicate a fairly constant effect of display across entry type, so no interactions were included in the model. An ANOVA was run against these subject, entry, and order, and the results shown in Table 12.

Table 12. ANOVA for vertical recall.

Analysis of Variance for VertScore, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	9.0667	9.0667	0.6476	2.38	0.009
Display	1	4.0111	4.0111	4.0111	14.77	0.000
Entry	2	0.2667	0.2667	0.1333	0.49	0.614
Error	72	19.5556	19.5556	0.2716		
Total	89	32.9000				

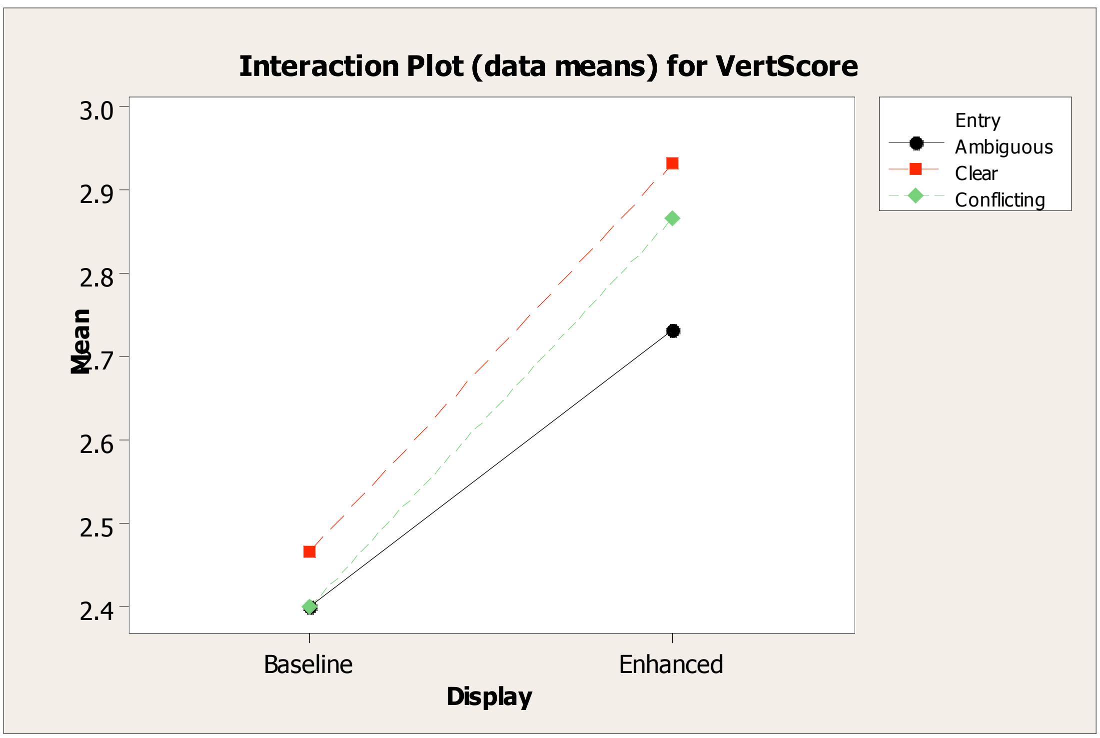


Figure 25. Interaction plot for vertical score between entry and display.

The ANOVA indicates a significant effect for subject (as expected) and display. The effect of display is shown graphically in Figure 26.

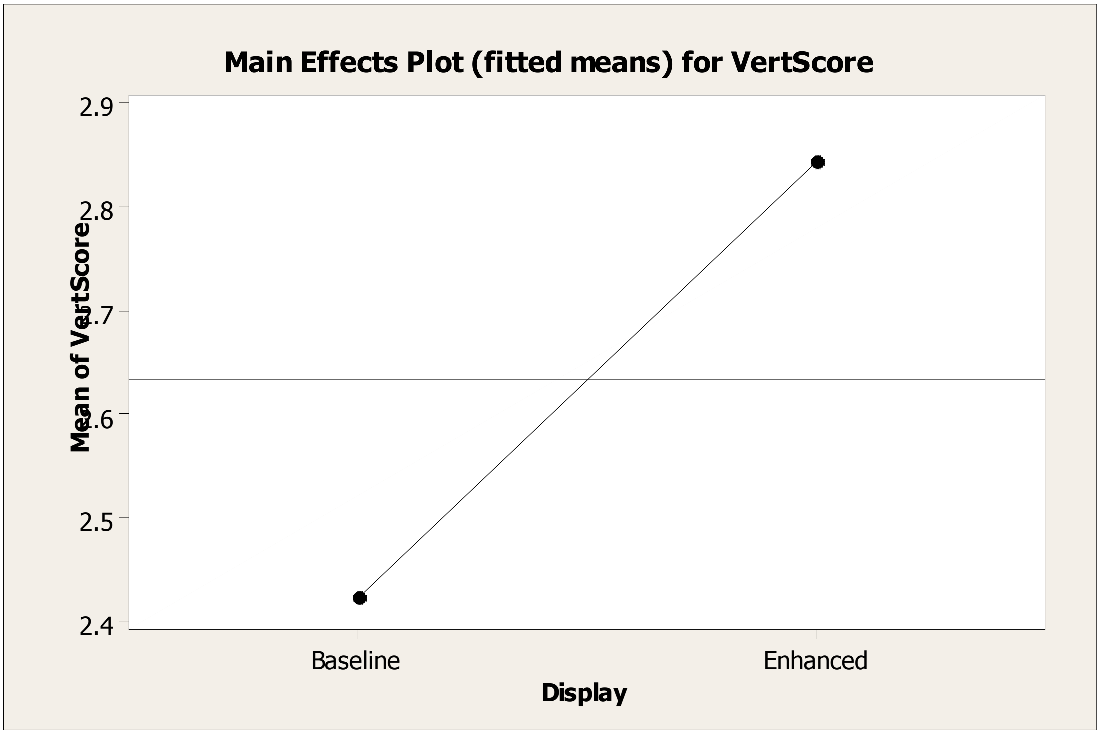


Figure 26. Main effects plot for other recall by subject and display.

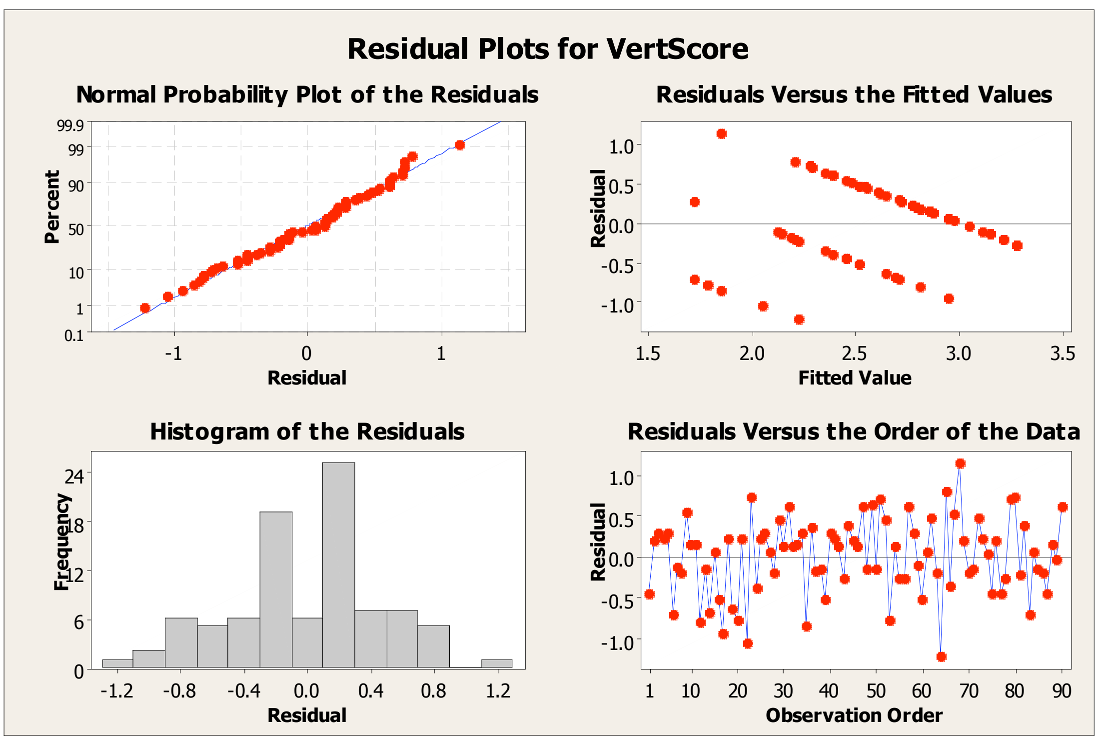


Figure 27. Residual plots for vertical score.

The residuals plots, shown in Figure 27, do not indicate that the residuals depart significantly from normality or constant variance, and also do not indicate any dependence on order of runs.

Recall of approach speed and missed approach deviations

Subjects were also questioned about approach speed and missed approach deviations. For these measures, the subject was asked one question each, so the results were either correct/incorrect. As categorical data, effects of the variables on the results were checked with nonparametric tests.

Approach speed

No significant effect of any variable was found on the ability of the subject to recall approach speed deviations.

Missed approach

No significant effect for subject, order, or entry was found to exist for the ability of the subject to recall missed approach deviations. However, a marginally significant effect (p=0.069) was indicated for display, as shown in Table 13.

Table 13. Nonparametric test for effect of display on recall of missed approach errors.

Kruskal-Wallis Test on MA

Display	N	Median	Ave Rank	Z
Baseline	45	1.000	40.5	-1.82
Enhanced	45	1.000	50.5	1.82
Overall	90		45.5	

H = 3.30 DF = 1 P = 0.069

A plot of the means for the missed approach recall by display is shown in Figure 28 (a correct answer was given a score of 1, an in correct answer a score of 0).

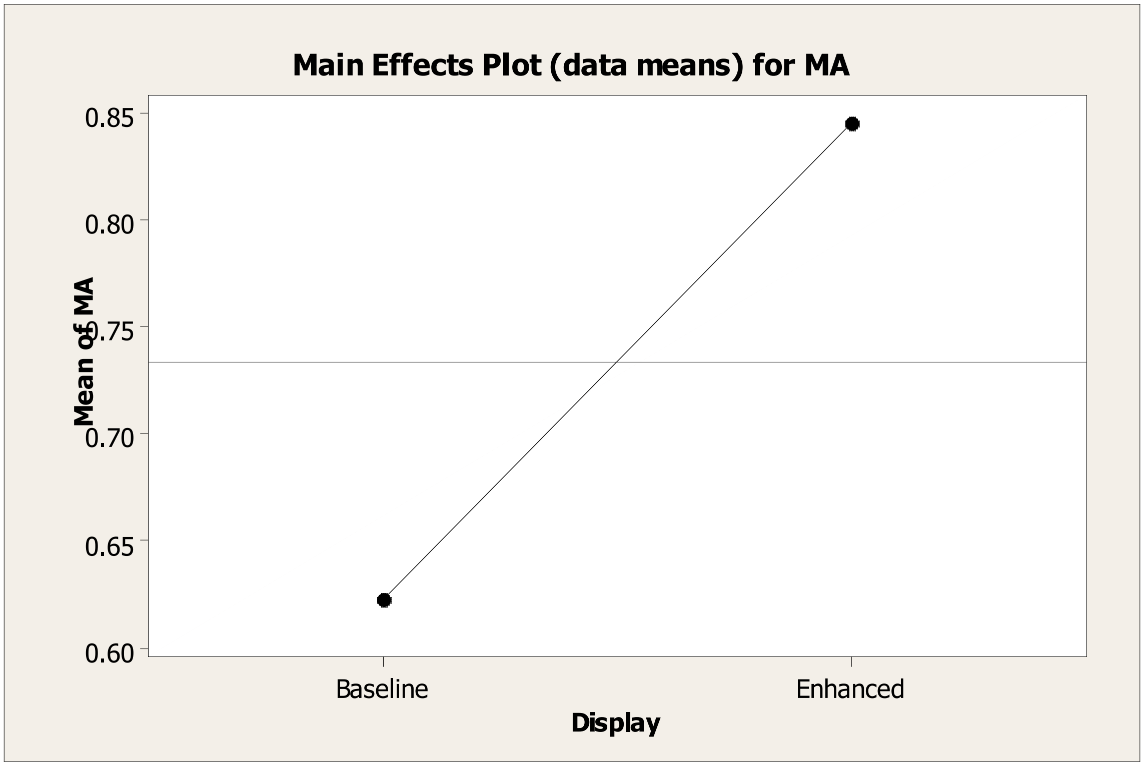


Figure 28. Plot of mean score for missed approach error recall by display.

NASA TLX workload ratings

Subjective assessments of workload ratings were taken upon completion of each experimental run. Workload was rated for 6 aspects: mental demand, physical demand, temporal demand, effort, performance, and frustration. The rating scale for each aspect allowed for a rating of between 0 and 24.

Mental_demand

Mental demand represents the amount of cognitive work that the subject expends to accomplish the task. The data summary for this measure is shown in Figure 29.

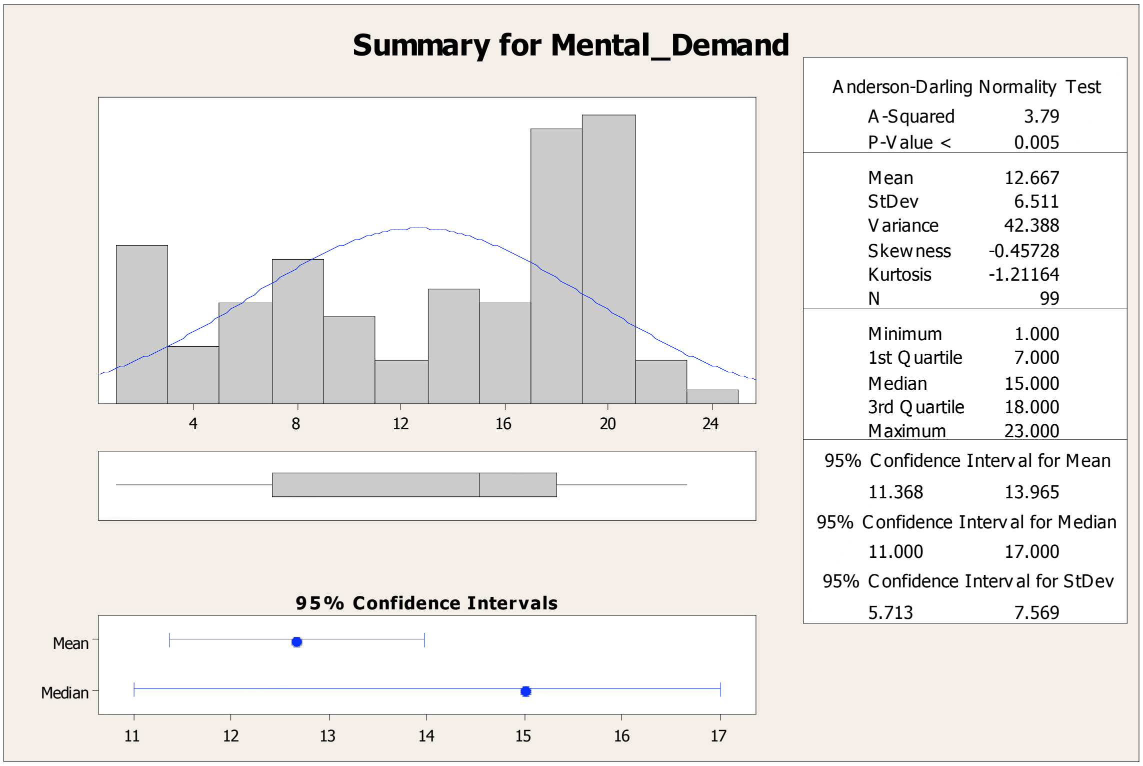


Figure 29. Mental demand summary.

This variable does not appear normal, and fails all tests of normality. Since the data seem to have multiple peaks, simple transformations will not bring it into normality. One might suspect that the distribution is being affected by significant differences in the base level of workload ratings for the subjects. In that case, the workload ratings could be reformatted as a "relative" workload (relative to the median workload).

As can be seen from the boxplots in Figure 30, this is the case. There seems to be a significant difference in not only the base level of workload rating, but also in the variance of workload across subjects.

The ratings were recalculated as the difference between the actual rating and the median rating for that subject. The resulting mental workload ratings are shown in Figure 31.

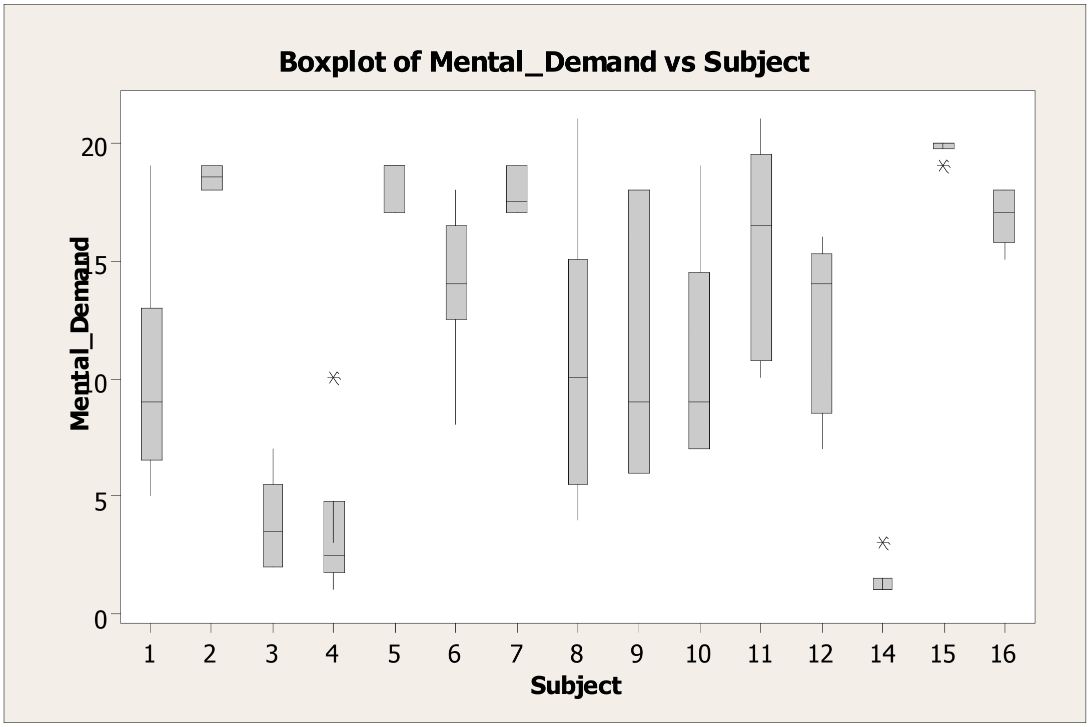


Figure 30. Boxplot of mental workload ratings for subjects.

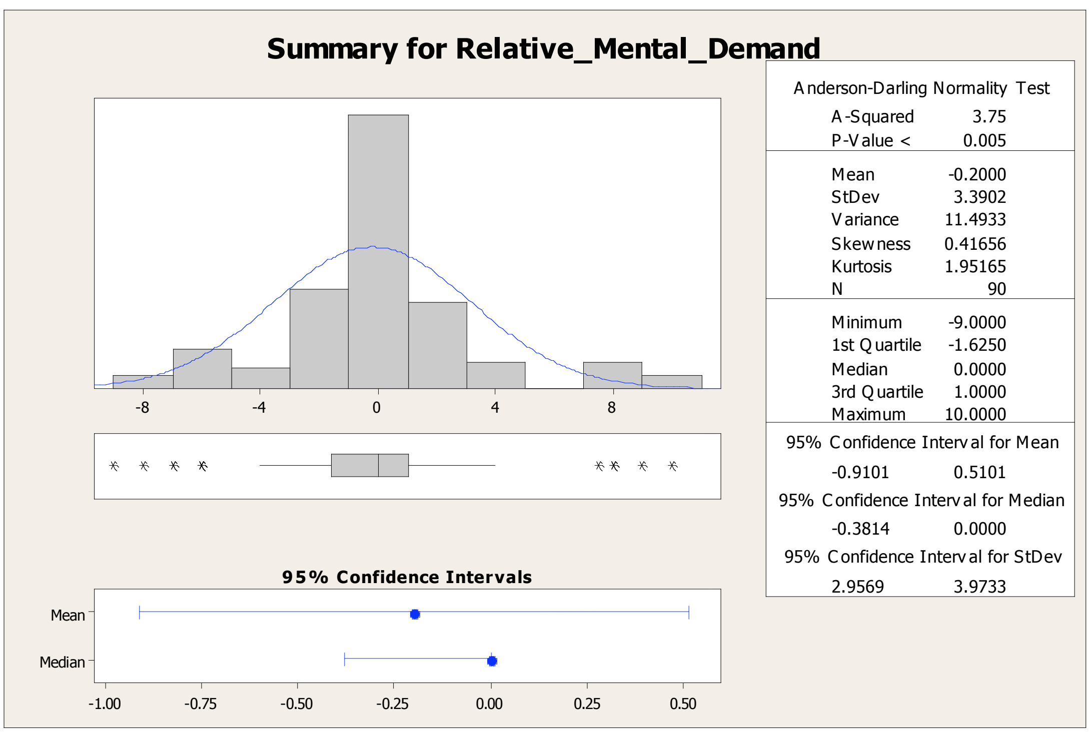


Figure 31. Graphical summary of relative mental workload ratings.

The data looks considerably more normal, but, due to its high kurtosis, it still fails all tests of normality. The histogram is reflecting the high number of times that the subject evaluated that no change in mental workload occurred between runs. Simple transformations were unable to alter this distribution sufficiently. Two approaches are considered and used here: (1) use ANOVA with a reduced _ (0.025) and (2) nonparametric f test (Kruskal-Wallis).

The results of the ANOVA are shown in Table 14. It indicates that there is only a significant effect of sequence of runs on mental workload. A plot of the means is shown in Figure 32, and shows that the earlier scenarios had higher workload than the later scenarios. This result is consistent with there being a learning effect on mental workload.

Table 14. Analysis of Variance for Relative_Mental_Demand, using Adjusted SS for Tests.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	102.400	102.400	7.314	0.73	0.734
Display	1	9.344	15.556	15.556	1.56	0.216
Entry	2	10.467	6.904	3.452	0.35	0.709
Sequence	5	232.081	232.081	46.416	4.65	0.001
Error	67	668.608	668.608	9.979		
Total	89	1022.900				

Tukey simultaneous tests on the above means indicates that the differences between 2 and 4, 5, and 6 are significant ($0.0172<p<0.0230$) and marginally significant between 1 and 4, 5, and 6 ($0.0429<p<0.0586$).

As mentioned above, Kruskal-Wallis nonparametric tests were run to confirm and check the results due to the high kurtosis of the data. The results are shown in Table 15(a) – (d).

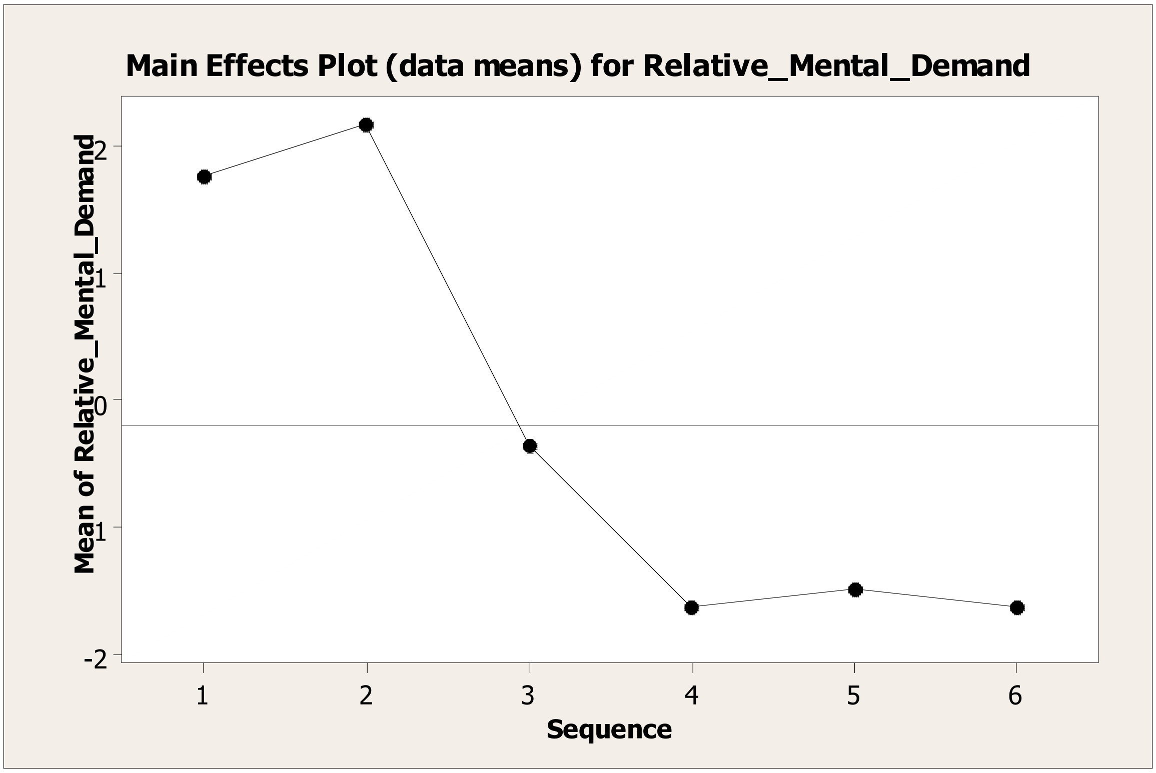


Figure 32. Means of mental demand workload vs. sequence of runs.

Table 15. Results of Kruskal-Wallis tests on effect of variables on mental workload.

```
(a)  Kruskal-Wallis Test for Subject

      H = 5.19  DF = 14  P = 0.983
      H = 5.33  DF = 14  P = 0.981  (adjusted for ties)

(b)  Kruskal-Wallis Test for Display

      H = 2.64  DF = 1  P = 0.104
      H = 2.71  DF = 1  P = 0.099  (adjusted for ties)

(c)  Kruskal-Wallis Test on for Entry

      H = 0.14  DF = 2  P = 0.934
      H = 0.14  DF = 2  P = 0.932  (adjusted for ties)

(d)  Kruskal-Wallis Test for Sequence

      H = 18.16  DF = 5  P = 0.003
      H = 18.65  DF = 5  P = 0.002  (adjusted for ties)
```

As expected, the test shows a significant effect of sequence. However, the test also shows a marginally significant effect for display, which was not indicated by the ANOVA. This discrepancy could occur if more early scenarios used the enhanced display. Pairwise comparisons were unable to differentiate between levels, but a plot of the means across display and sequence (Figure 33) shows that the means are lower for the enhanced display for all runs except the first and fifth runs, where the means are nearly equal. The largest differences in means occur for the second, third, and sixth runs, where the means for the enhanced displays are lower.

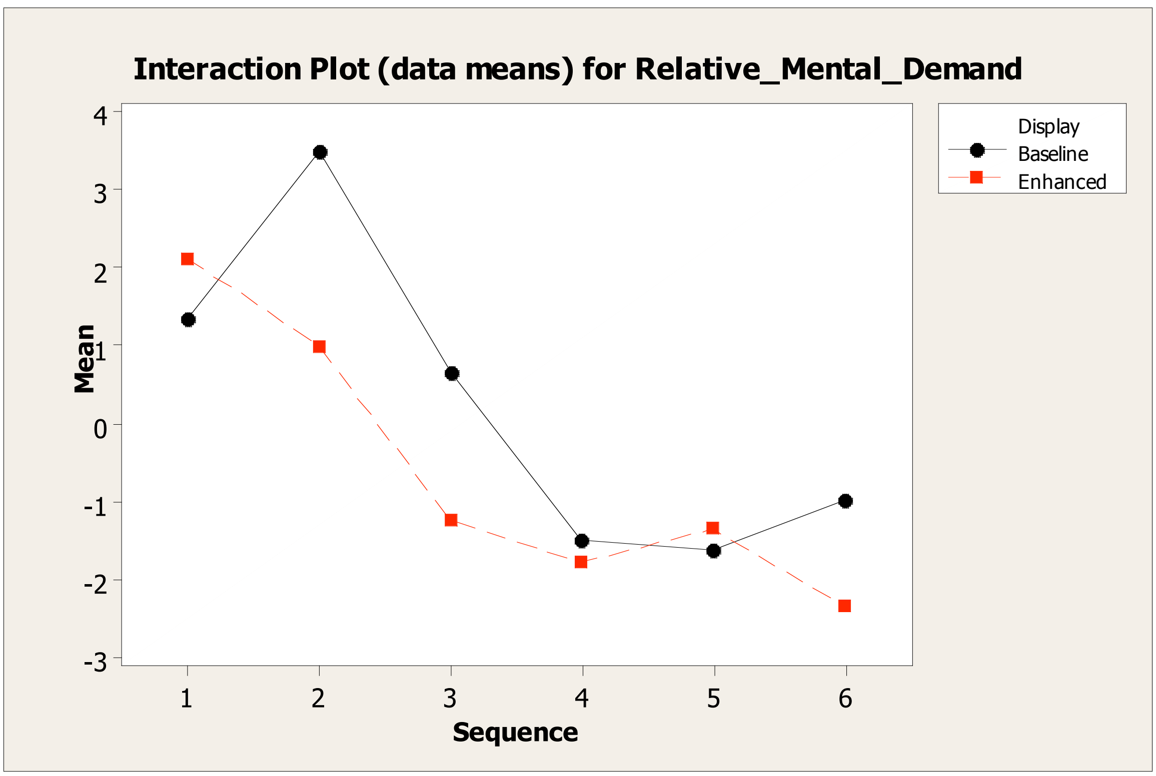


Figure 33. Plot of means of mental workload for sequence and display.

Physical_demand

The ratings for physical demand were also normalized across each subject to provide a relative physical demand measure for each scenario. A summary of the data (Figure 34) shows that this distribution is again not normal and the same technique used for mental workload will be used.

An ANOVA was run and the results shown in Table 16. It again shows a significant effect for sequence.

Table 16. Analysis of Variance for Relative_Physical_Demand, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	135.622	135.622	9.687	1.13	0.350
Display	1	0.044	1.029	1.029	0.12	0.730
Entry	2	16.022	16.573	8.287	0.97	0.386
Sequence	5	155.224	155.224	31.045	3.62	0.006
Error	67	574.709	574.709	8.578		
Total	89	881.622				

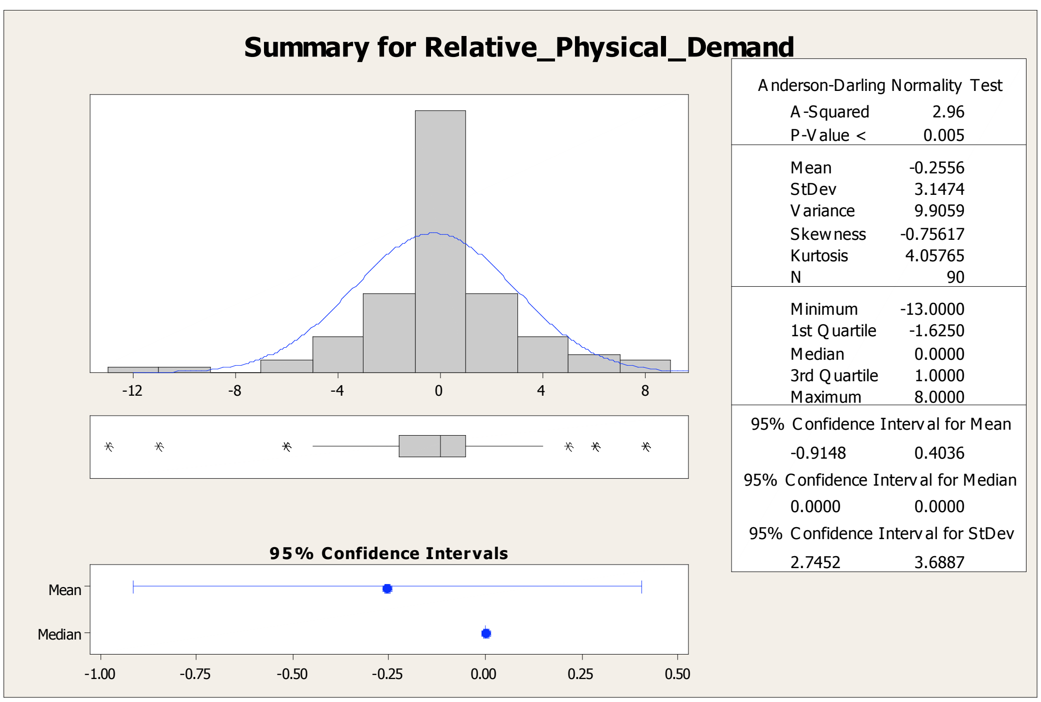


Figure 34. Graphical summary of physical demand workload measure.

A plot of the means for physical demand against sequence is shown in Figure 35, and, as expected, shows lower workload measures for later runs. Kruskal-Wallis tests were run to support the ANOVA results and are shown in Table 17(a) to (d).

Table 17. Nonparametric test results for physical workload measure.

```
(a)  Kruskal-Wallis Test for Subject

       H = 6.48  DF = 14  P = 0.953
       H = 6.63  DF = 14  P = 0.948  (adjusted for ties)

(b)  Kruskal-Wallis Test for Display

       H = 0.01  DF = 1  P = 0.926
       H = 0.01  DF = 1  P = 0.925  (adjusted for ties)

(c)  Kruskal-Wallis Test on for Entry

       H = 1.13  DF = 2  P = 0.569
       H = 1.15  DF = 2  P = 0.561  (adjusted for ties)

(d)  Kruskal-Wallis Test for Sequence

       H = 20.88  DF = 5  P = 0.001
       H = 21.38  DF = 5  P = 0.001  (adjusted for ties)
```

Tukey simultaneous tests on the above means indicates that the difference between 2 and 5 is significant (p=0.0253) and marginally significant between 1 and 5 (p=0.0873) and between 2 and 6 (p=0.0949).

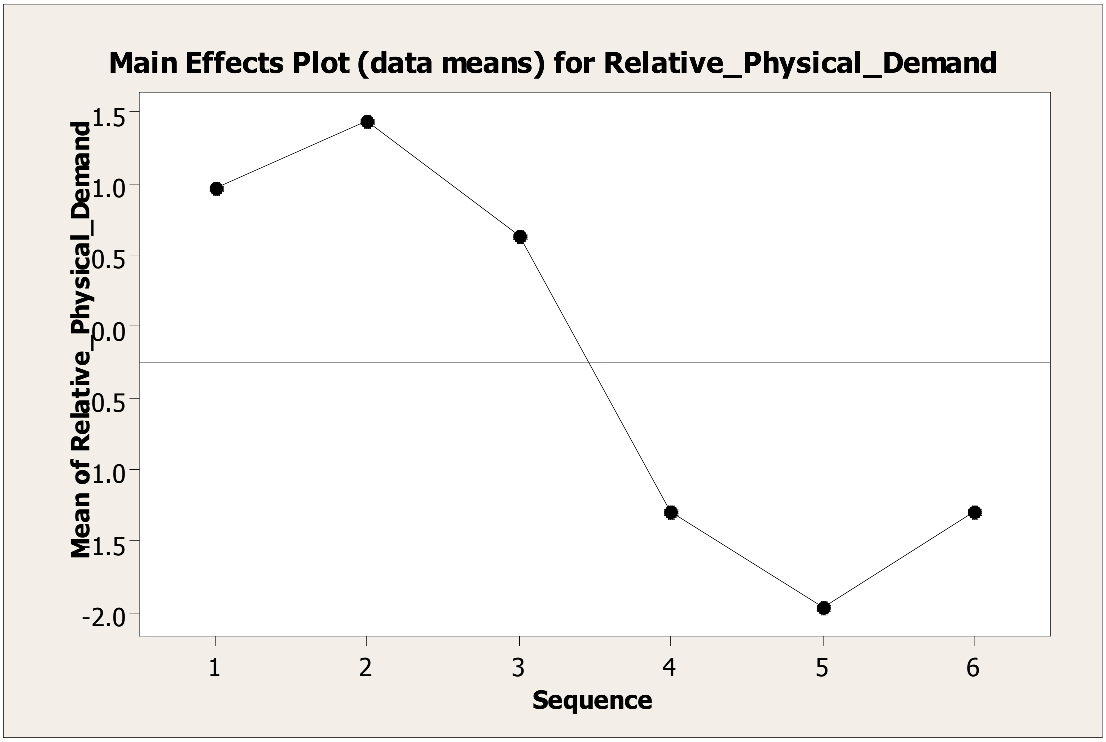


Figure 35. Plot of means of mental demand workload rating vs. sequence.

<u>Temporal_demand</u>

As previously, the ratings for temporal demand were normalized across each subject to provide a relative temporal demand measure for each scenario. A summary of the data (Figure 36) shows that this distribution is again not normal and the same technique used for mental workload will be used.

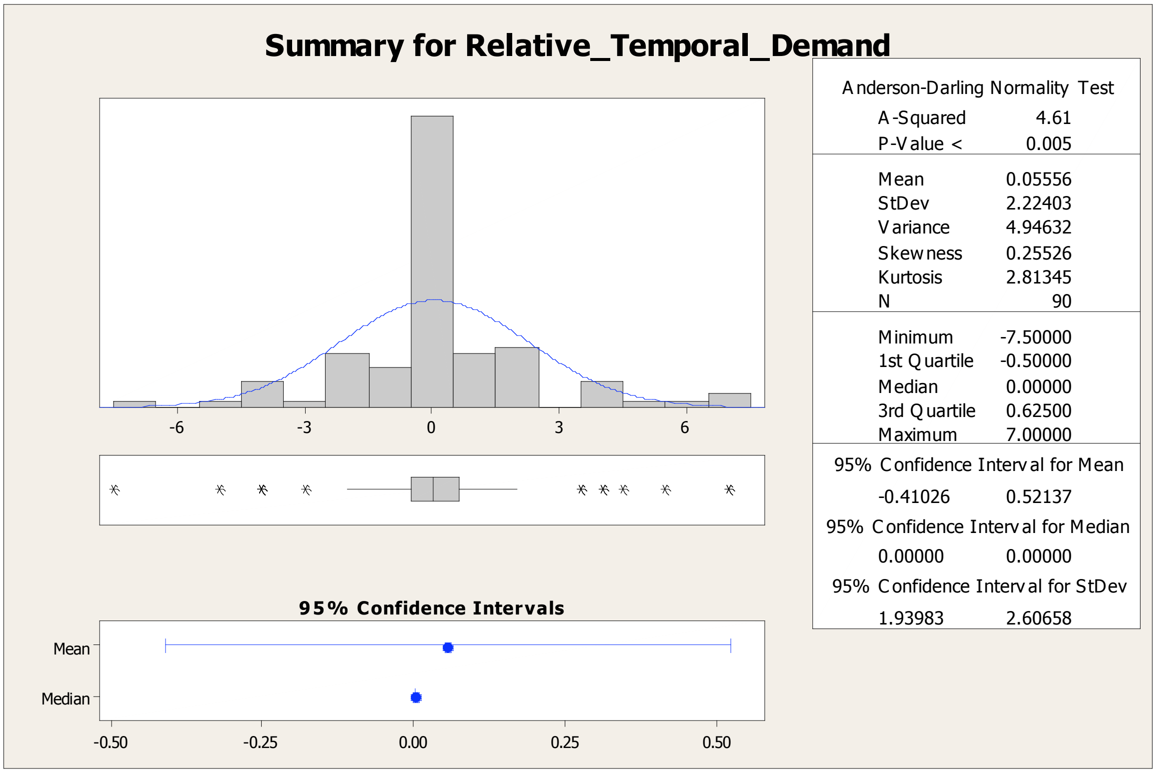


Figure 36. Graphical summary of the temporal demand workload ratings.

An ANOVA was run and the results shown in Table 18. It shows a marginally significant effect for sequence only.

Table 18. Analysis of Variance for Relative_Temporal_Demand, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	36.556	36.556	2.611	0.51	0.916
Display	1	2.844	4.346	4.346	0.86	0.358
Entry	2	2.156	2.556	1.278	0.25	0.778
Sequence	5	58.924	58.924	11.785	2.32	0.052
Error	67	339.742	339.742	5.071		
Total	89	440.222				

A plot of the means for physical demand against sequence is shown in Figure 37, and, as expected, shows lower workload measures for later runs. Kruskal-Wallis tests were run to support the ANOVA results and are shown in Table 19(a) to (d).

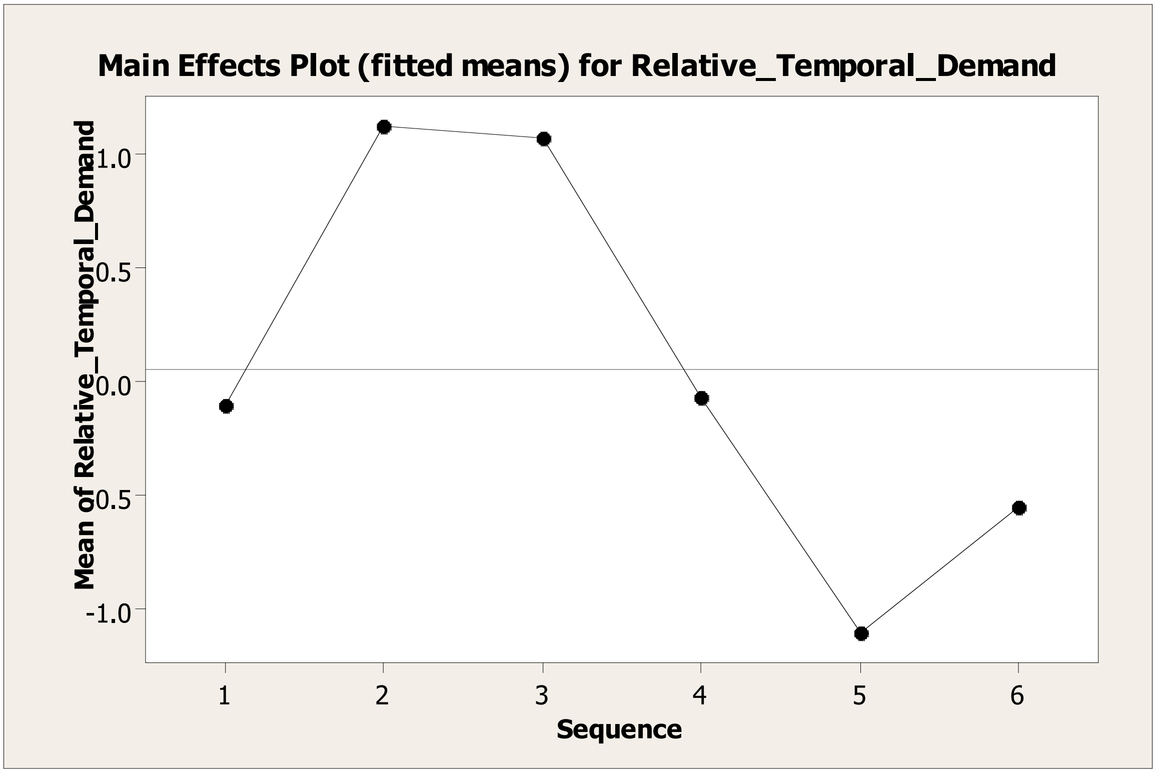


Figure 37. Plot of means of tmeporal demand workload rating vs. sequence.

Table 19. Nonparametric test results for physical workload measure.

```
(a)  Kruskal-Wallis Test for Subject

       H = 7.73  DF = 14  P = 0.903
       H = 8.43  DF = 14  P = 0.866  (adjusted for ties

(b)  Kruskal-Wallis Test for Display

       H = 2.01  DF = 1  P = 0.157
       H = 2.19  DF = 1  P = 0.139  (adjusted for ties)

(c)  Kruskal-Wallis Test on for Entry

       H = 0.10  DF = 2  P = 0.951
       H = 0.11  DF = 2  P = 0.946  (adjusted for ties)


(d)  Kruskal-Wallis Test for Sequence

       H = 14.09  DF = 5  P = 0.015
       H = 15.36  DF = 5  P = 0.009  (adjusted for ties)
```

Tukey simultaneous tests on the above means indicates that a marginally significant difference could be detected between 2 and 5 (p = 0.0867) only.

Effort

As previously, the ratings for effort were normalized across each subject to provide a relative effort measure for each scenario. A summary of the data (Figure 38) shows a better distribution. The distribution fails the Anderson-Darling normality test, but passes both the Kolmogorov-Smirnov and Ryan-Joiner normality tests. Since the distribution does not appear to depart significantly from normality, only parametric tests will be used.

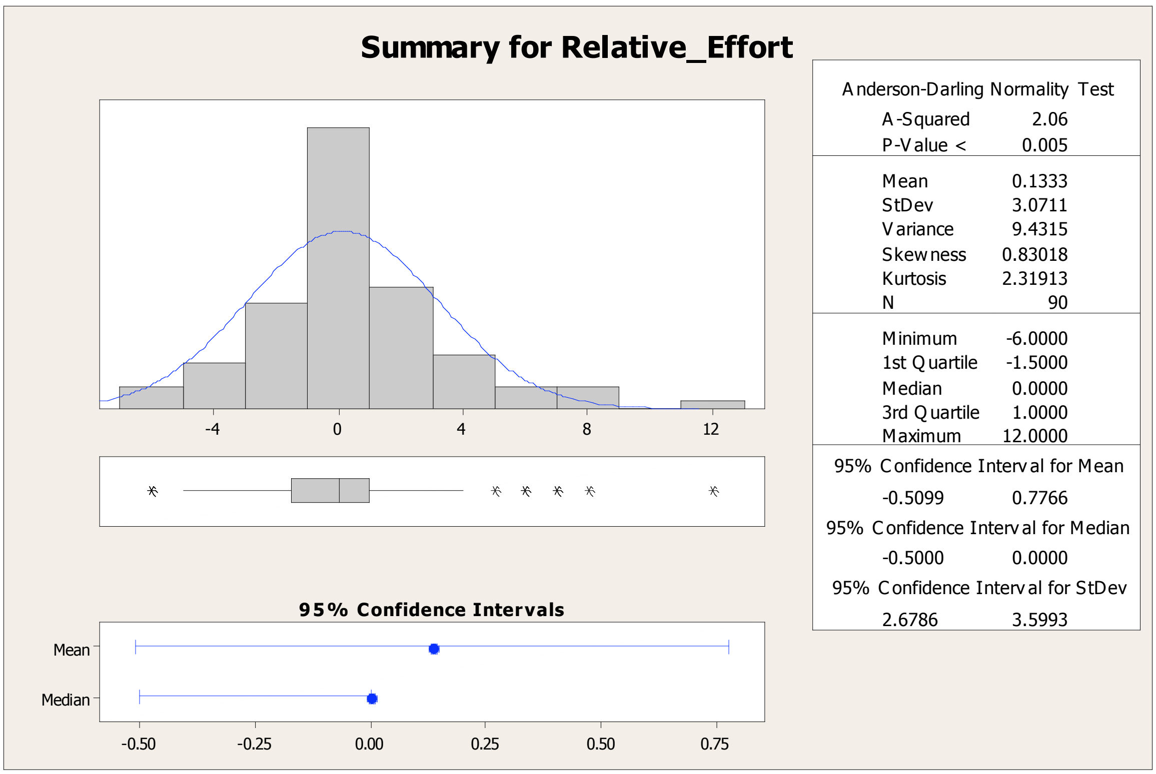


Figure 38. Graphical summary of effort rating.

An ANOVA was run and the results shown in Table 20. It shows a significant effect for both subject and sequence.

Table 20. Analysis of Variance for Effort, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	2673.07	2673.07	190.93	19.74	0.000
Display	1	0.40	2.10	2.10	0.22	0.642
Entry	2	2.07	2.14	1.07	0.11	0.895
Sequence	5	154.71	154.71	30.94	3.20	0.012
Error	67	648.15	648.15	9.67		
Total	89	3478.40				

A plot of the means for physical demand against sequence is shown in Figure 39, and, as expected, shows lower workload measures for later runs.

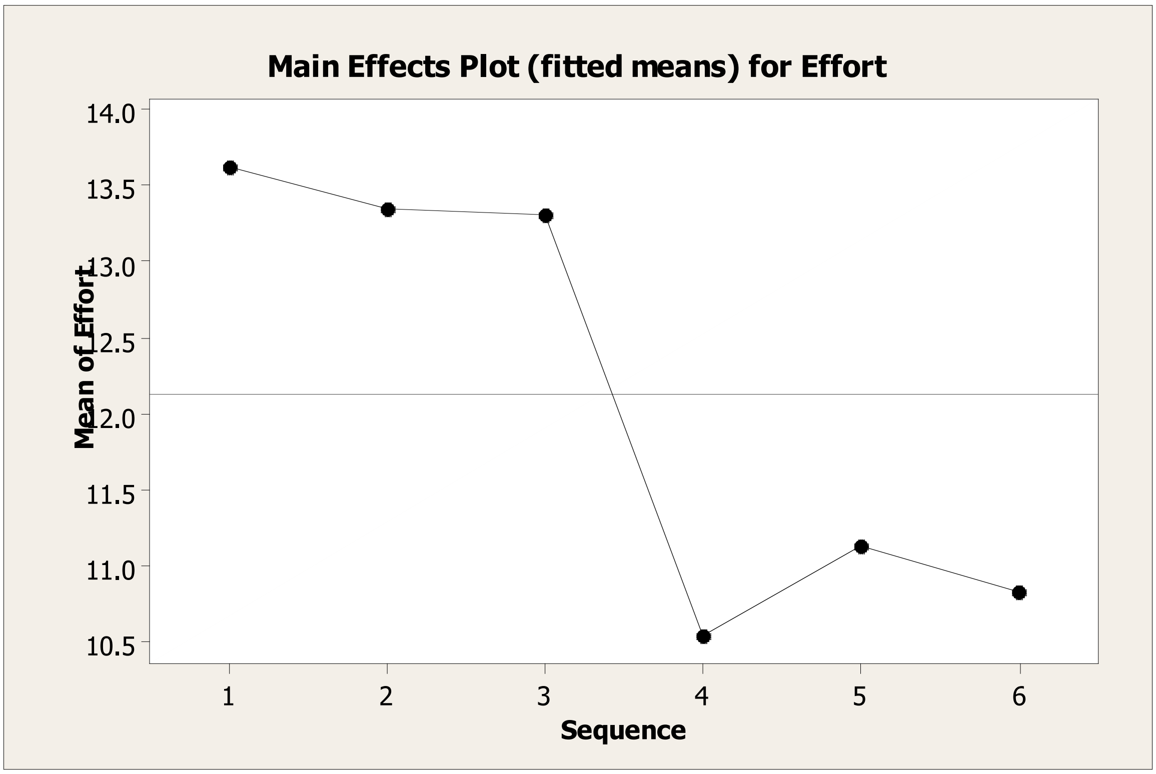


Figure 39. Plot of means of effort rating vs. sequence.

Tukey simultaneous tests on the above means indicates that a marginally significant difference could be detected between 1 and 4 ($p = 0.0833$) only.

Figure 40 shows the means for each subject. Deviation from zero for a subject indicates a difference between the median and mean, since these ratings were relative to the median. This occurs when one or more of the subject's ratings are relatively high or low. Figure 41(a) shows the ratings for subject 8, and Figure 39(b) shows the ratings for subject 12.

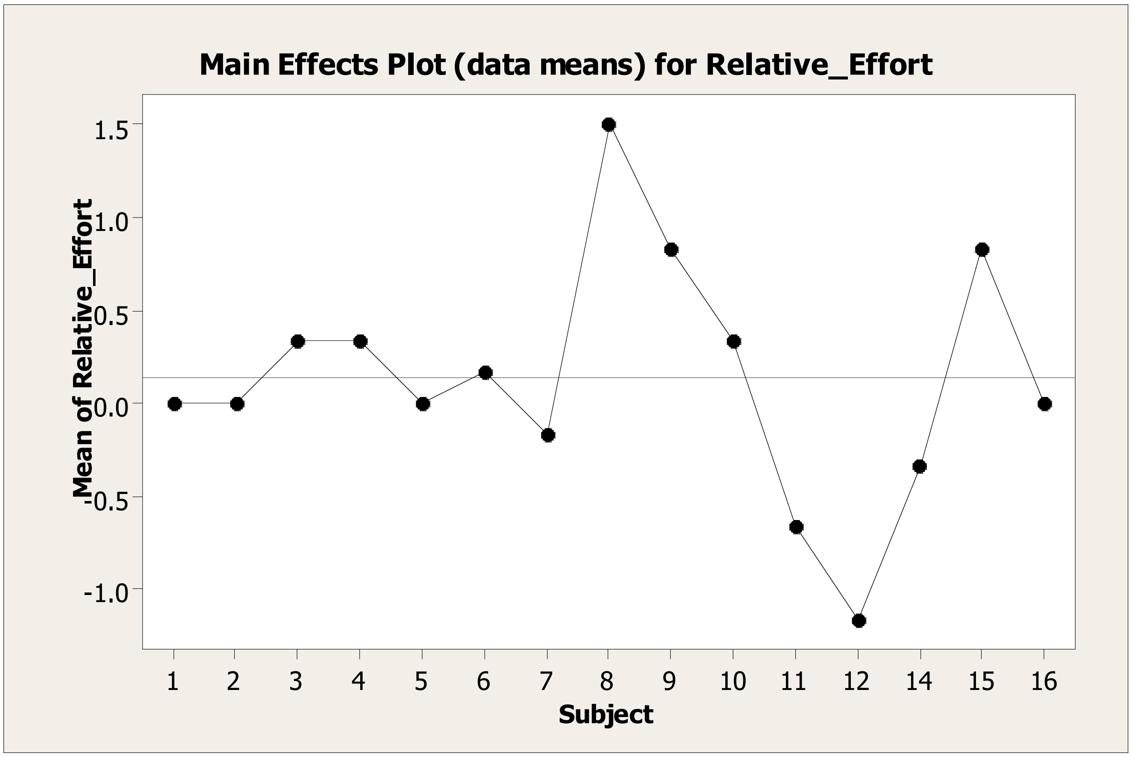
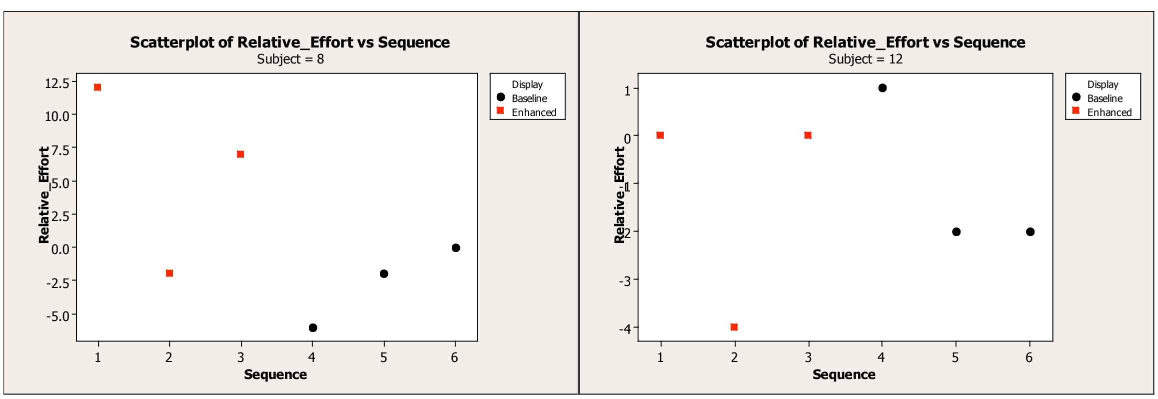


Figure 40. Plot of means of effort ratings by subject.



(a) (b)

Figure 41. Plot of effort measures for subjects 8 and 12.

Subject 8 rated the first and third runs (both of which used the enhanced display) much higher than the median. Subject 12 rated the second run (which used the enhanced display) well below the median.

Performance

As previously, the ratings for performance were normalized across each subject to provide a relative performance measure for each scenario. A summary of the data (Figure 42) shows that this data is again not normal and the same technique used for mental workload will be used here.

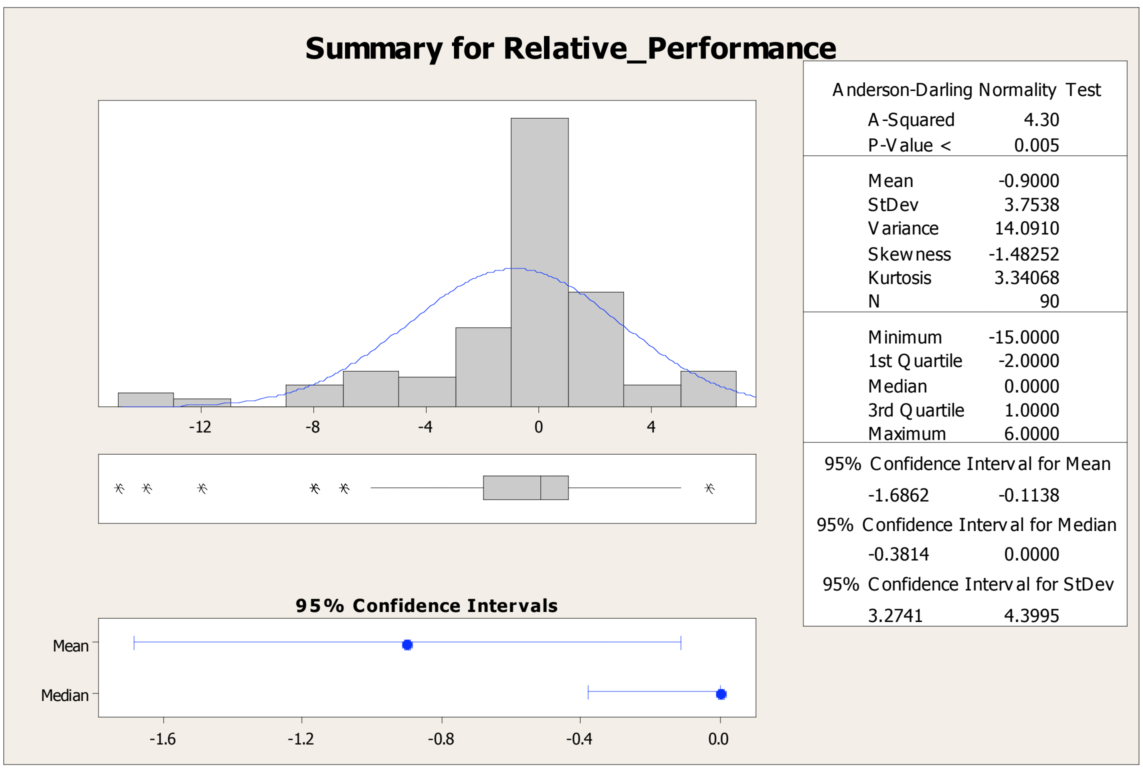


Figure 42. Graphical summary of performance ratings.

An ANOVA was run and the results shown in Table 21. It shows a marginally significant effect for sequence only.

Table 21. Analysis of Variance for Relative_Performance, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	146.60	146.60	10.47	0.77	0.697
Display	1	1.34	3.10	3.10	0.23	0.635
Entry	2	17.27	15.28	7.64	0.56	0.573
Sequence	5	177.86	177.86	35.57	2.62	0.032
Error	67	911.02	911.02	13.60		
Total	89	1254.10				

A plot of the means for performance against sequence is shown in Figure 43, and, as expected, shows lower workload measures for later runs. Kruskal-Wallis tests were run to support the ANOVA results and are shown in Table 22 (a) to (d).

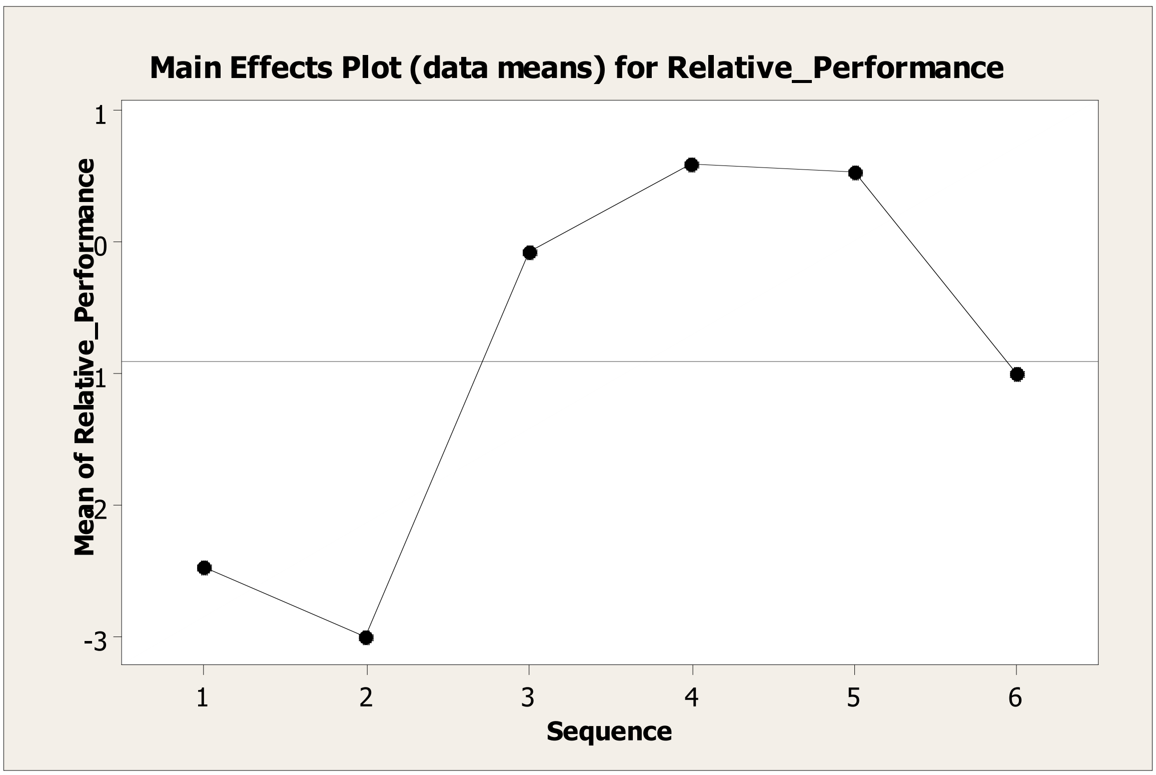


Figure 43. Plot of means of performance rating vs. sequence.

Tukey simultaneous tests on the above means found no detectable differences by sequence.

Table 22. Nonparametric test results for performance measure.

```
(a)  Kruskal-Wallis Test for Subject

       H = 5.17  DF = 14  P = 0.983
       H = 5.31  DF = 14  P = 0.981  (adjusted for ties)

(b)  Kruskal-Wallis Test for Display

       H = 0.03  DF = 1  P = 0.869
       H = 0.03  DF = 1  P = 0.867  (adjusted for ties)

(c)  Kruskal-Wallis Test on for Entry

       H = 3.09  DF = 2  P = 0.213
       H = 3.17  DF = 2  P = 0.205  (adjusted for ties)

(d)  Kruskal-Wallis Test for Sequence

       H = 10.11  DF = 5  P = 0.072
       H = 10.39  DF = 5  P = 0.065  (adjusted for ties)
```

Frustration

As previously, the ratings for frustration were normalized across each subject to provide a relative frustration measure for each scenario. A summary of the data (Figure 44) shows that this distribution is again not normal and the same technique used for mental workload will be used here. It also shows a long right tail, indicating a few very high frustration ratings relative to the median.

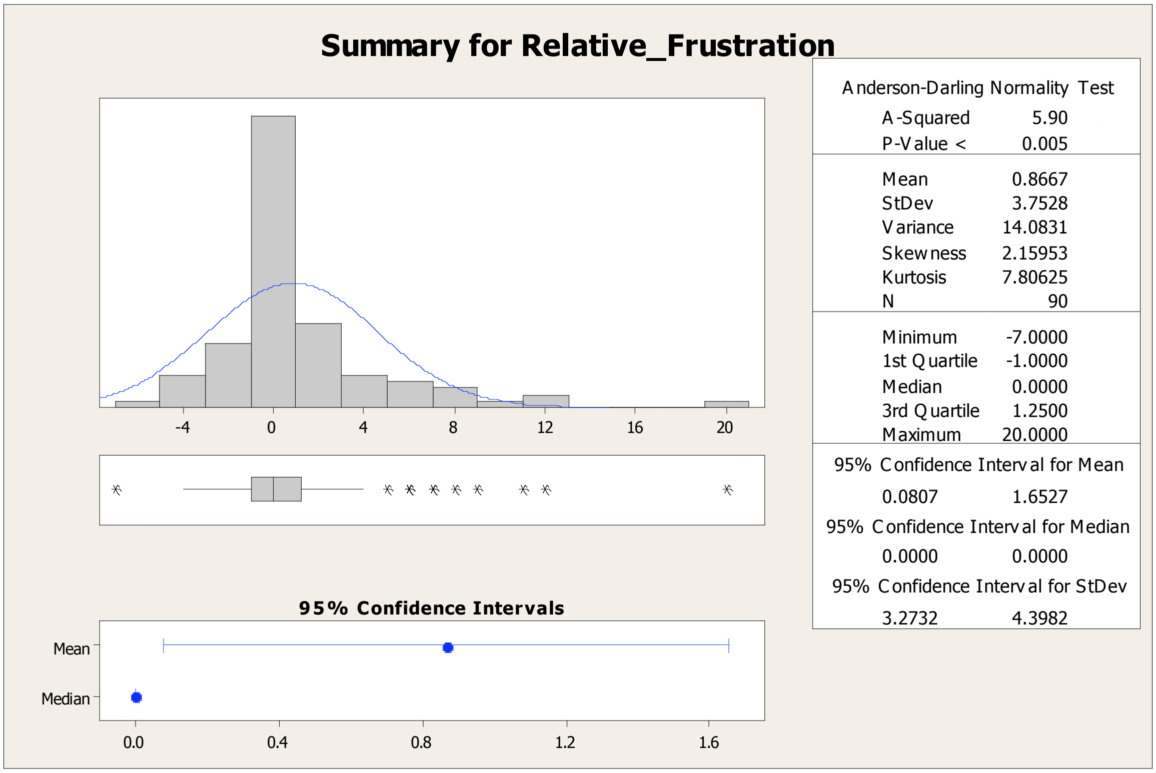


Figure 44. Graphical summary for frustration ratings.

An ANOVA was run and the results shown in Table 23. It shows a marginally significant effect for sequence only.

Table 23. Analysis of Variance for Relative_Frustration, using Adjusted SS for Tests.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	156.40	156.40	11.17	0.90	0.567
Display	1	0.04	0.40	0.40	0.03	0.858
Entry	2	11.47	11.29	5.64	0.45	0.638
Sequence	5	250.05	250.05	50.01	4.01	0.003
Error	67	835.44	835.44	12.47		
Total	89	1253.40				

A plot of the means for performance against sequence is shown in Figure 45, and, as expected, shows lower workload measures for later runs. Kruskal-Wallis tests were run to support the ANOVA results and are shown in Table 24 (a) to (d).

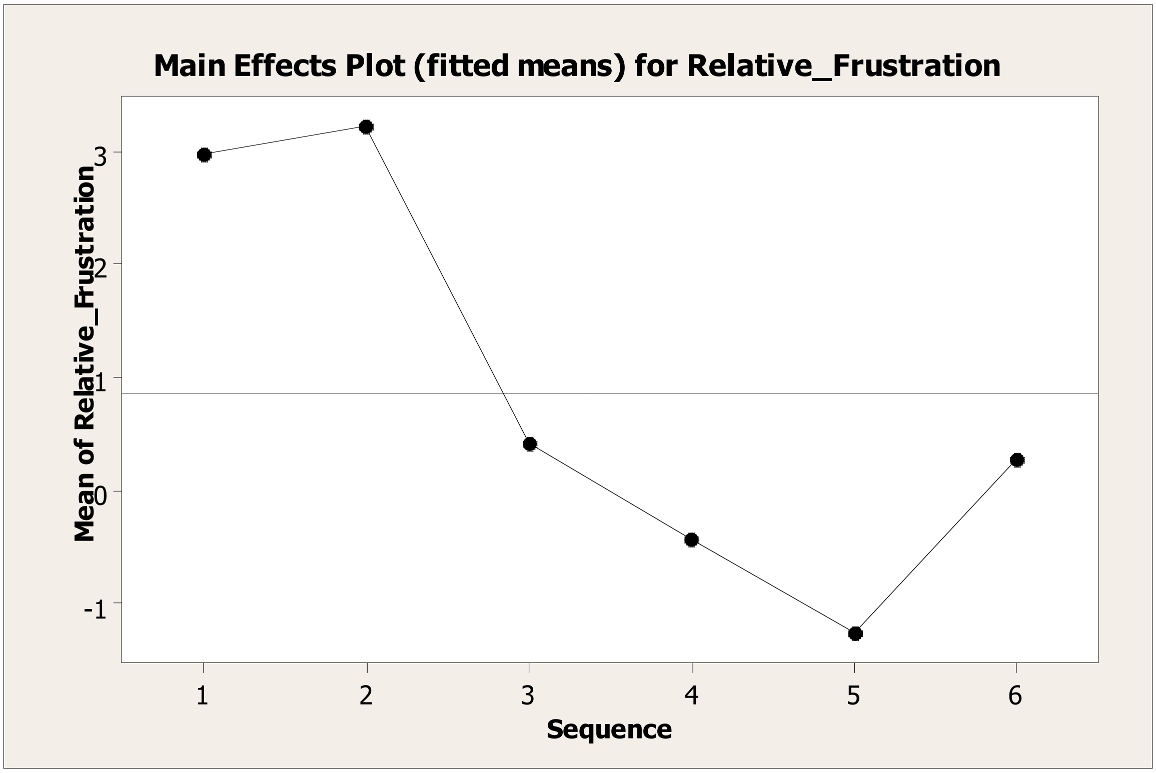


Figure 45. Plot of frustration rating means vs. sequence.

Table 24. Nonparametric test results for performance measure.

```
(a)   Kruskal-Wallis Test for Subject

        H = 5.10  DF = 14  P = 0.984
        H = 5.28  DF = 14  P = 0.981  (adjusted for ties)

(b)   Kruskal-Wallis Test for Display

        H = 0.05  DF = 1  P = 0.815
        H = 0.06  DF = 1  P = 0.812  (adjusted for ties)

(c)   Kruskal-Wallis Test on for Entry

        H = 1.66  DF = 2  P = 0.436
        H = 1.72  DF = 2  P = 0.423  (adjusted for ties)

(d)   Kruskal-Wallis Test for Sequence

        H = 20.45  DF = 5  P = 0.001
        H = 21.19  DF = 5  P = 0.001  (adjusted for ties)
        )
```

Tukey simultaneous tests on the above means found significant differences between run 1 and 5 (p=0.0200) and run 2 and 5 (p=0.0114), and a marginally significant difference between runs 2 and 4 (p=0.0673).

Number, frequency, and magnitude of control inputs

An analysis was conducted on the subject's control inputs. The subjects' inputs were accomplished using the yoke, which affected simulated control movements of the ailerons (for bank control) and elevator (for pitch control). These control inputs were caused changes in the bank angle and the pitch angle, which were also recorded. As the latter are more meaningful and exactly correlated with control movements, they will be used for the analysis.

The measures were grouped over 30 second periods and a mean and variance were calculated. The mean would then indicate the general trend of the control and the variance would be a measure of the stability of the control inputs. As such, the mean is unimportant, but high variance would indicate instability and is undesirable, particularly on the latter parts of the approach.

Bank angle

A data summary for the bank is shown in Figure 46.

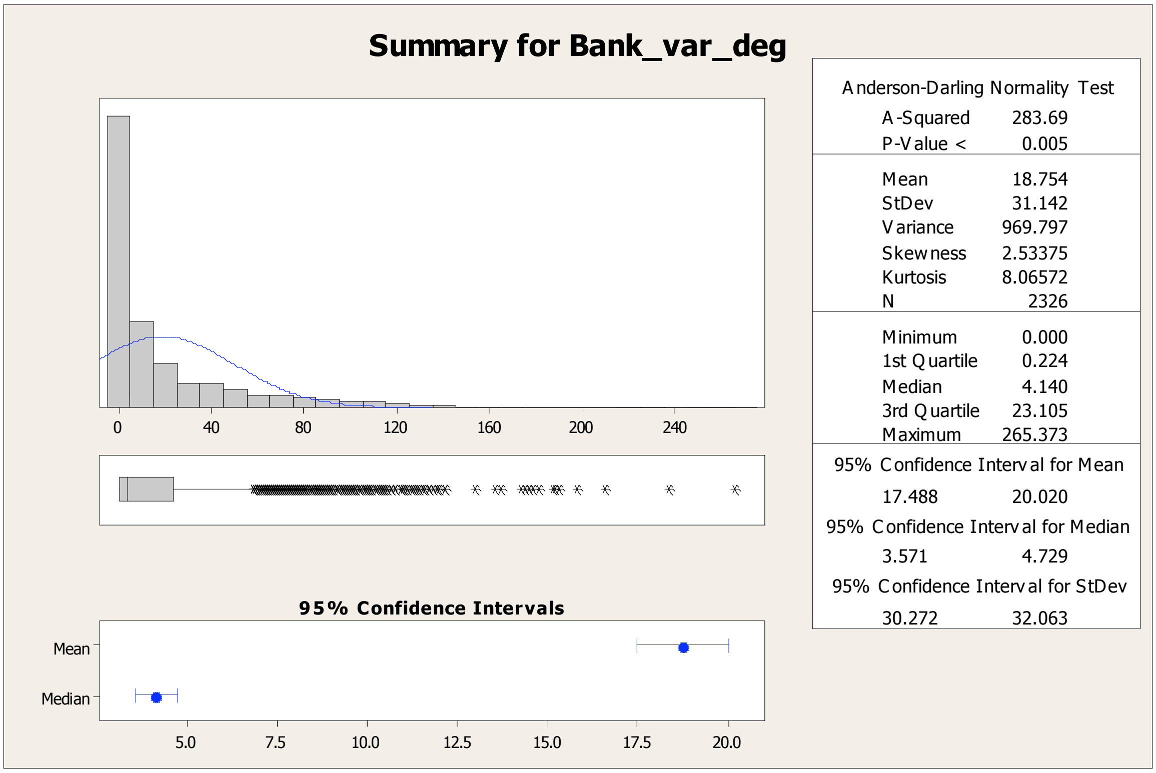


Figure 46. Data summary for bank variance.

The summary shows that the distribution is not normal, with the majority of variance measures being near zero with a number of very high outliers. No outliers could be eliminated, and transformations were unsuccessful at normalizing the data.

Boxplots for the bank variance (Figure 47) shows differences in variance for different periods and for different subjects. Nonparametric tests showed significant differences between subject ($p<0.001$), order ($p = 0.02$), and time ($p<0.001$), but failed to find any significant difference for entry or display.

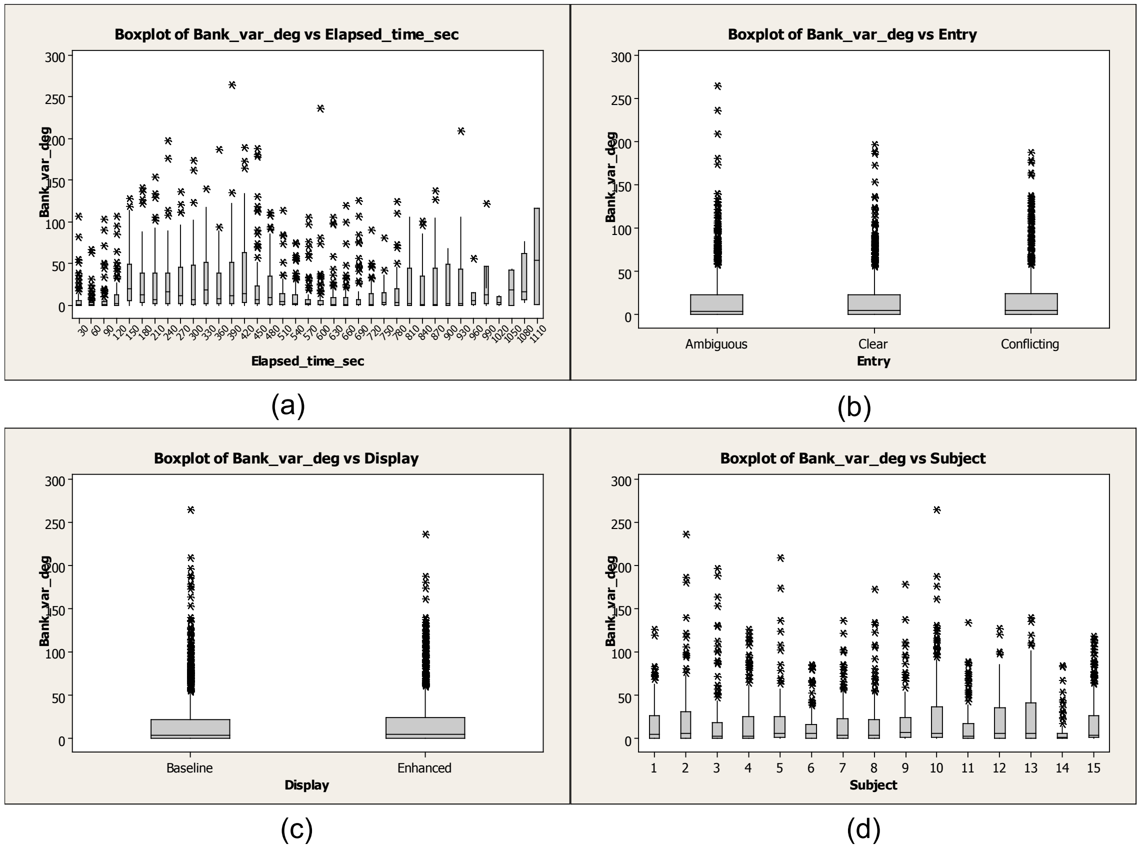


Figure 47. Boxplots of bank variance by time, entry, display, and subject.

<u>Pitch</u>

The data summary for pitch is shown in Figure 48. The data is very similar to the bank angle data, and again is not normal. No outliers could be removed, and transformations failed to normalize the data. Boxplots for time, subject, entry, and display are shown in Figure 49. Nonparametric tests again indicated significant difference for elapsed time (p<0.001), subject (p<0.001), and order of runs (p=0.016).

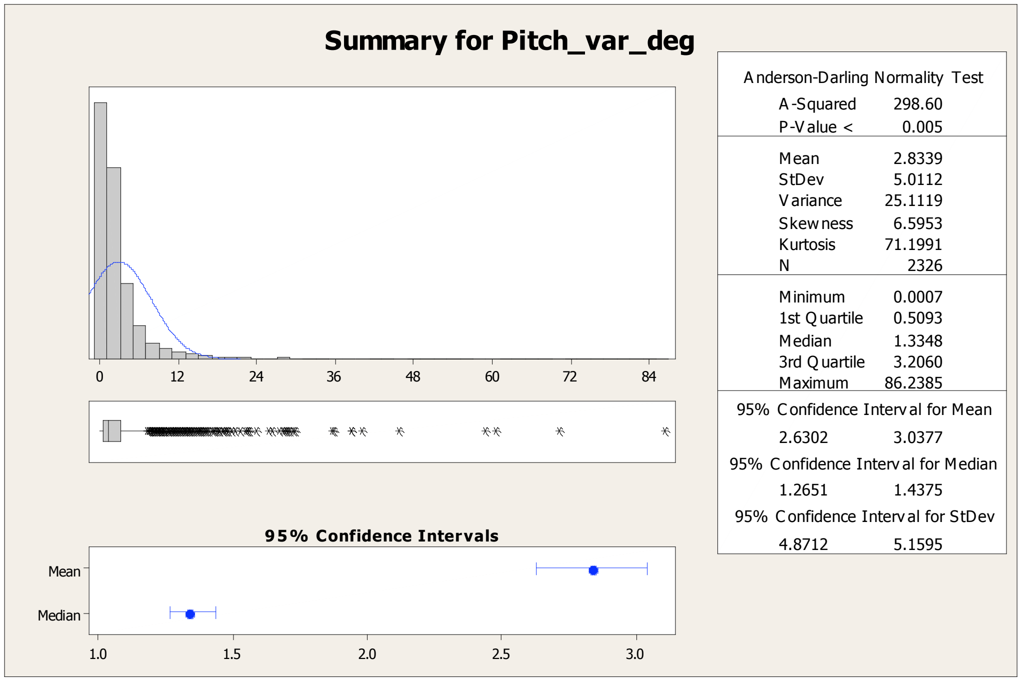


Figure 48. Data summary for pitch variance.

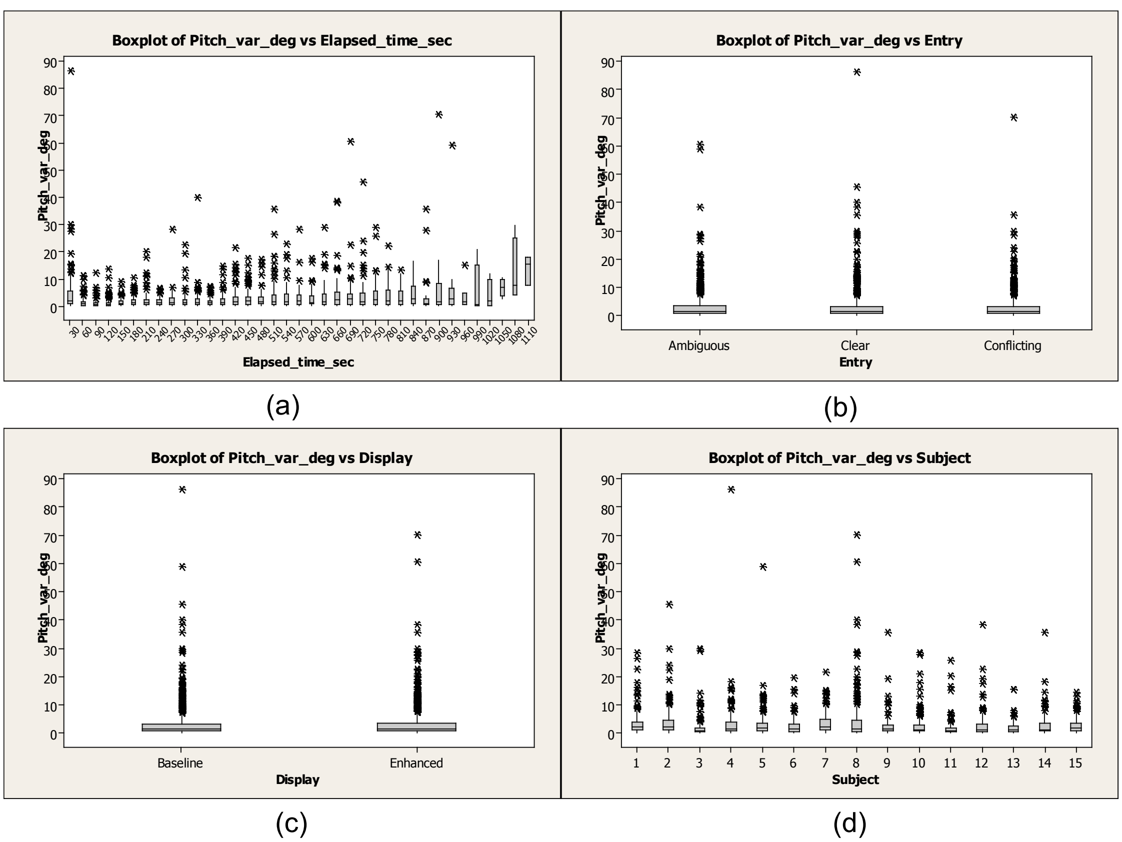


Figure 49. Boxplots of pitch variance for elapsed time, entry, display, and subject.

Throttle

The variance of the throttle movements were also recorded in the same fashion as for pitch and bank. A graphical summary of the data is shown in Figure 50. The data is obviously not normal, with high kurtosis and large outliers. The boxplot in Figure 51 shows that, as with pitch and bank, a large amount of the variance occurs at the later parts of the approach. No outliers could be removed, and transformations were unable to normalize the data.

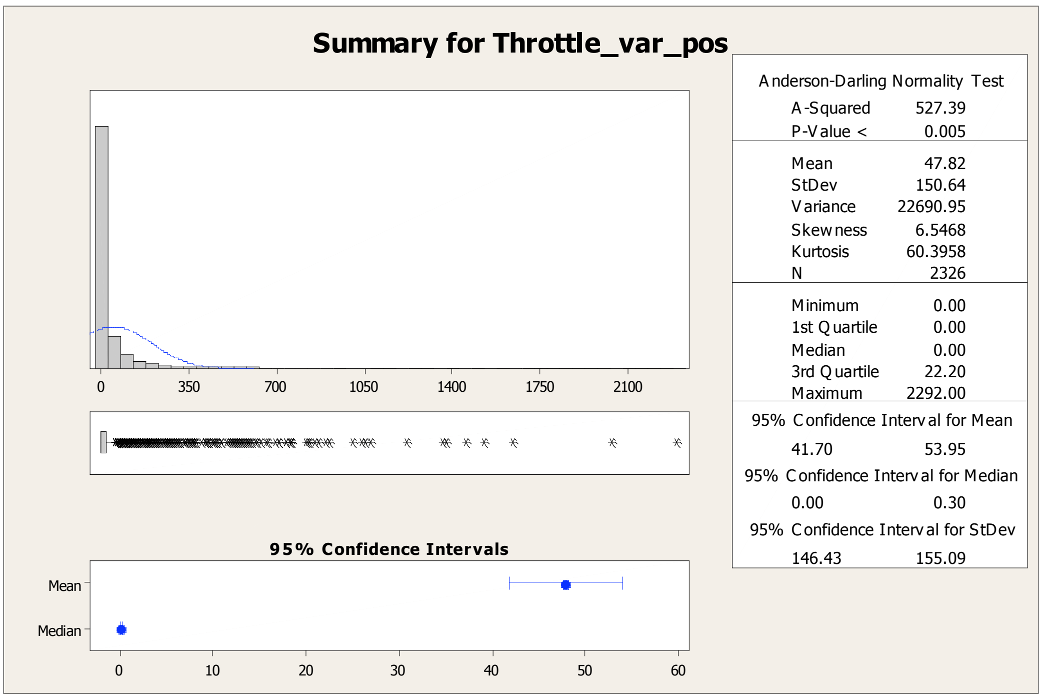


Figure 50. Data summary for throttle movements.

Nonparametric tests showed significance by subject ($p<0.001$) and elapsed time ($p<0.001$).

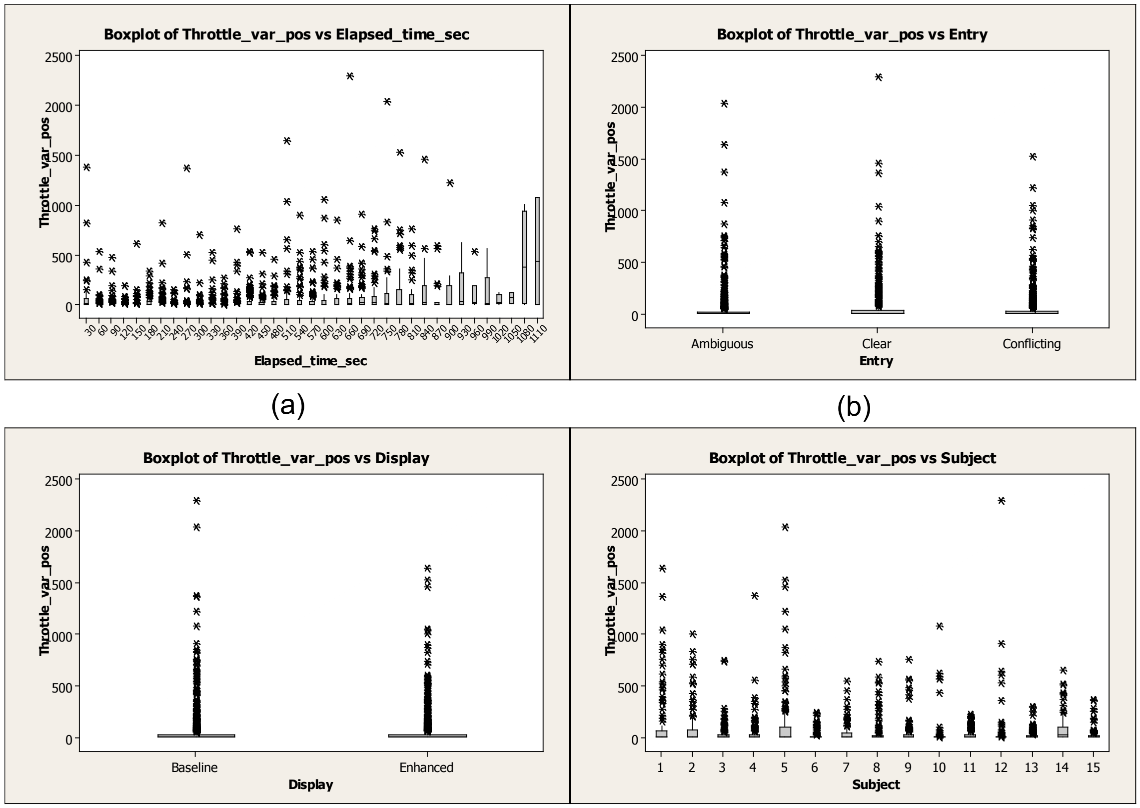


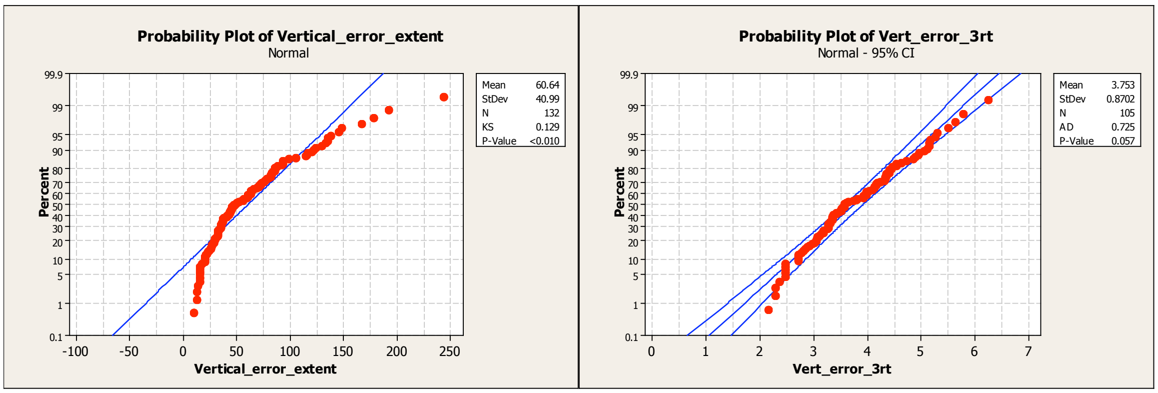
Figure 51. Boxplot of throttle movement by time, entry, display, and subject.

Summary statistics of flight error

An analysis of flight errors was accomplished for both the lateral and vertical flight paths. The flight paths of the runs was examined and compared against the required lateral flight path and the required vertical flight path. The extent and duration of each deviation was recorded, along with the portion of the approach on which the error occurred.

<u>Altitude errors</u>

A probability plot of the vertical extent error (Figure 52a) shows that the distribution is not normal, but a third root transformation normalizes the distribution (Figure 52b).



(a) (b)

Figure 52. Probability plots of vertical error and third root transformation.

Figure 53 shows that interactions may exist between the display and entry variable, so this was included in the model. An ANOVA was then run on the third root of the vertical error extent measure, with the results shown in Table 25.

Table 25. Analysis of Variance for Vertical_error_3rt, using Adjusted SS for Tests

Analysis of Variance for Vertical_error_3rt, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	14	28.6579	28.2712	2.0194	3.73	0.000
Display	1	0.0002	0.0014	0.0014	0.00	0.959
Entry	2	0.1347	0.1926	0.0963	0.18	0.837
Display*Entry	2	0.2742	0.2742	0.1371	0.25	0.777
Error	112	60.7092	60.7092	0.5420		
Total	131	89.7762				

The residuals plots (Figure 54) show no violation of assumptions of constant error variance or normality of error terms. The results show only a significant effect by subject.

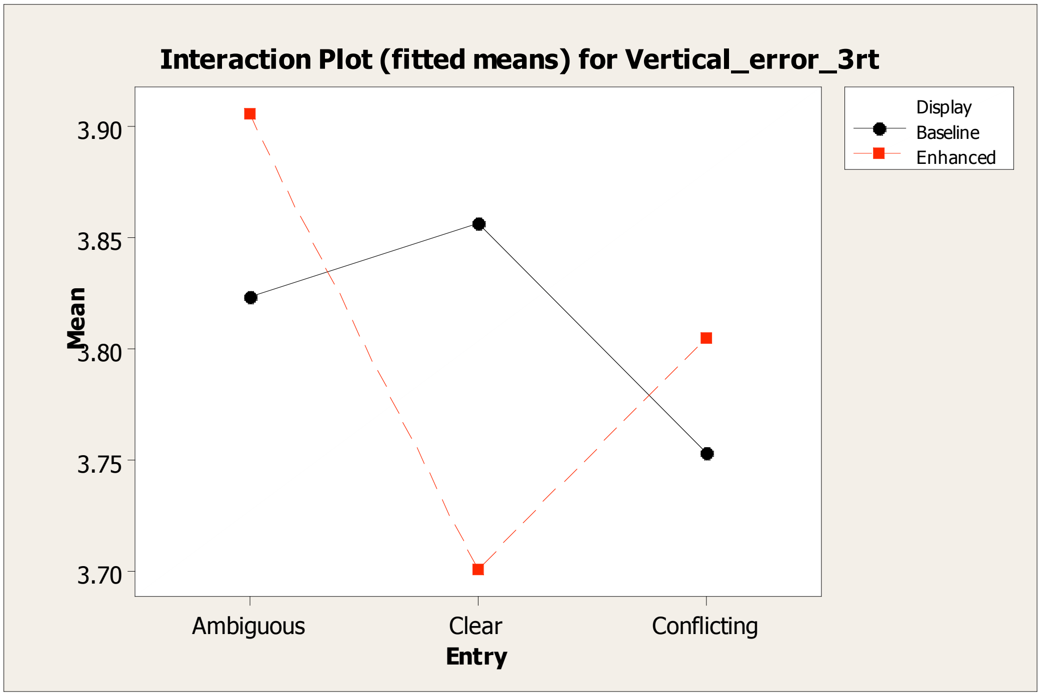


Figure 53. Interaction plot between entry and display for third root of vertical error.

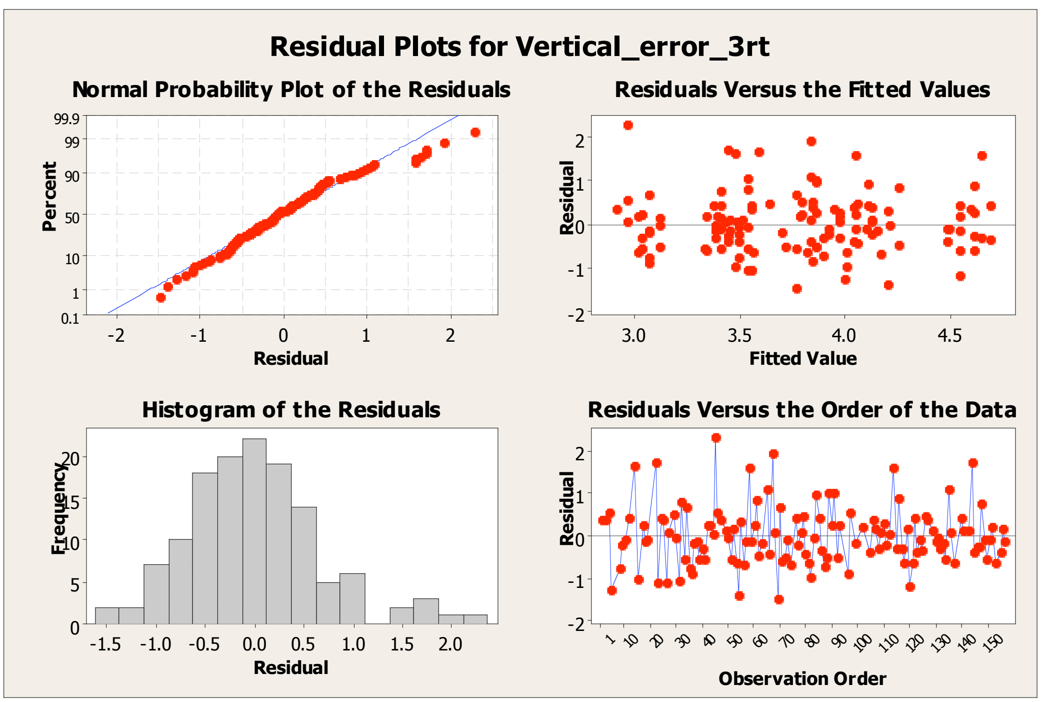


Figure 54. Residuals plots for third root of vertical error extent.

Next, the altitude error duration was examined. Figure 55 shows that the distribution is not normal, and Figure 56 shows that interactions exist between the display variable and the location of error variable. To eliminate this interaction, the errors at each location will be examined separately.

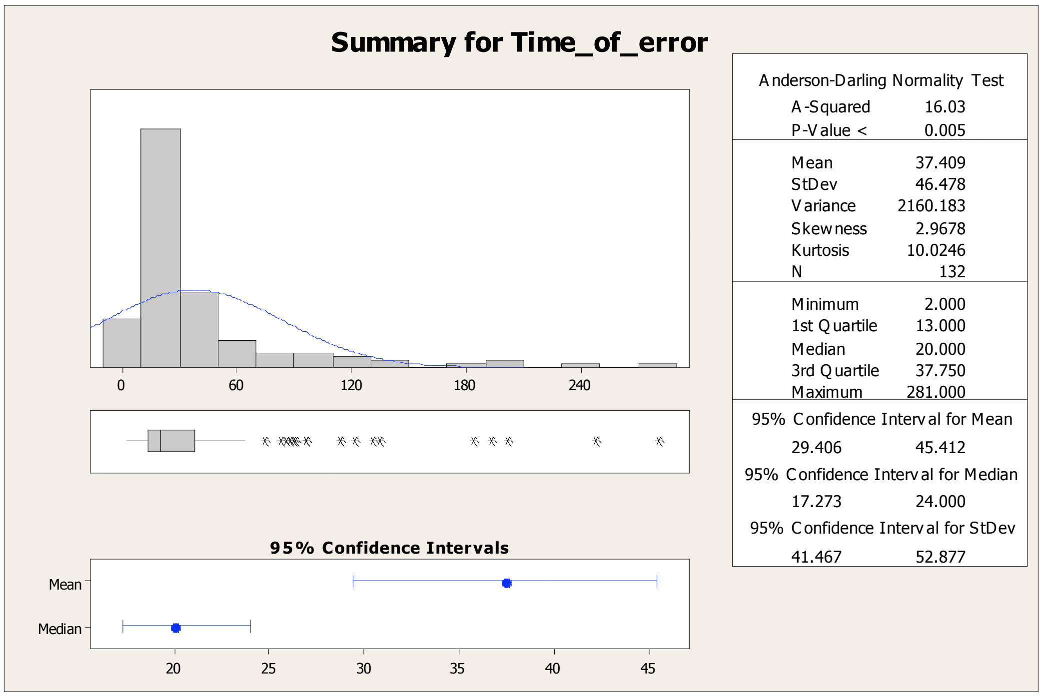


Figure 55. Data summary for altitude error durations.

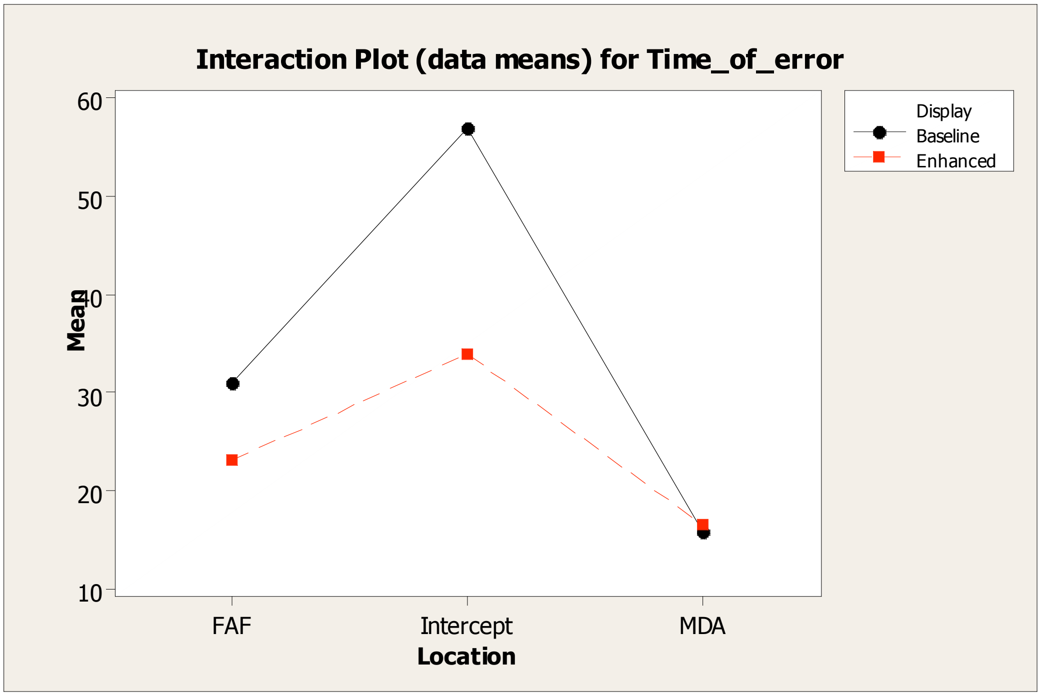


Figure 56. Interaction plot of altitude error duration vs. location of error.

For the intercept location, the data was non-normal and transformations failed to achieve normality. Nonparametric tests (Kruskal-Wallis) detected significant difference between subjects (p=0.012) and marginally significant results for display (0.077) and entry (0.078).

For the FAF, the data was non-normal, but a square root transformation normalized the data. An ANOVA was run, with the results shown in Table 26 but failed to detect any significant differences. The residuals plots (Figure 57) again confirm normality and constant variance of error terms.

Table 26. ANOVA for square root of error duration at FAF.

```
Analysis of Variance for Time_Sqrt, using Adjusted SS for Tests

Source          DF   Seq SS   Adj SS   Adj MS     F      P
Order            1    0.016    0.014    0.014  0.01  0.942
Subject         10   51.662   43.354    4.335  1.72  0.195
Display          1    0.577    0.451    0.451  0.18  0.681
Entry            2    1.748    1.730    0.865  0.34  0.717
Display*Entry    2    1.542    1.542    0.771  0.30  0.743
Error           11   27.806   27.806    2.528
Total           27   83.350
```

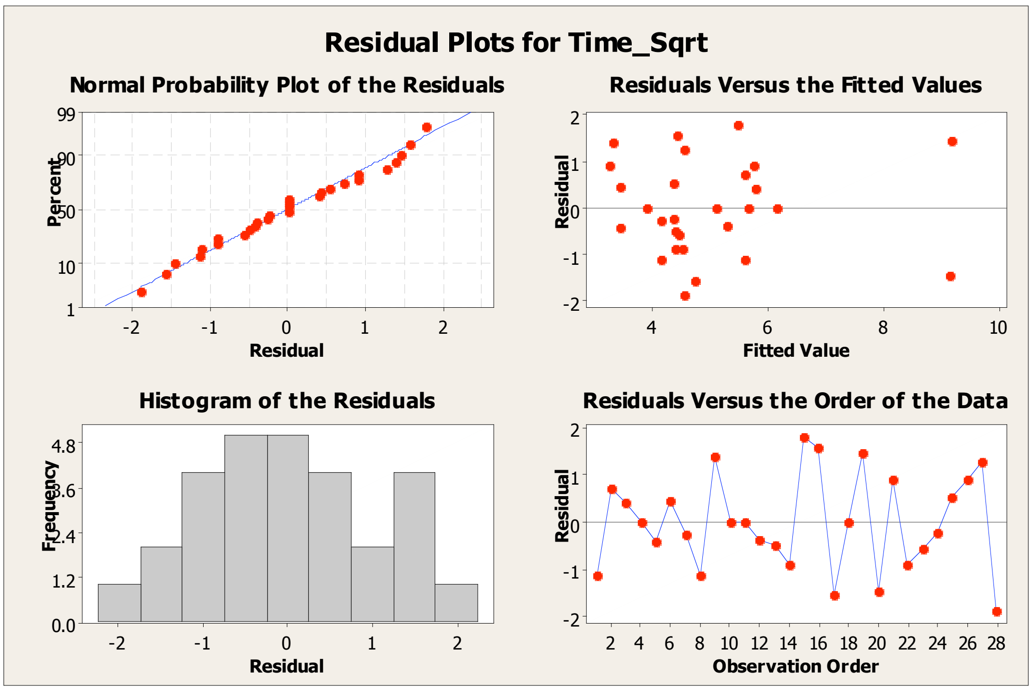


Figure 57. Residuals plots for square root of vertical error duration at FAF.

For the MDA, the square root transformation again normalized the data. The ANOVA results shown in Table 27 failed to detect any significant differences. The residuals plots in Figure 58 confirm the assumptions of normalcy and constant variance of error terms.

Table 27. ANOVA for square root of vertical error duration at MDA.

Analysis of Variance for Time_Sqrt, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Order	1	0.016	0.014	0.014	0.01	0.942
Subject	10	51.662	43.354	4.335	1.72	0.195
Display	1	0.577	0.451	0.451	0.18	0.681
Entry	2	1.748	1.730	0.865	0.34	0.717
Display*Entry	2	1.542	1.542	0.771	0.30	0.743
Error	11	27.806	27.806	2.528		
Total	27	83.350				

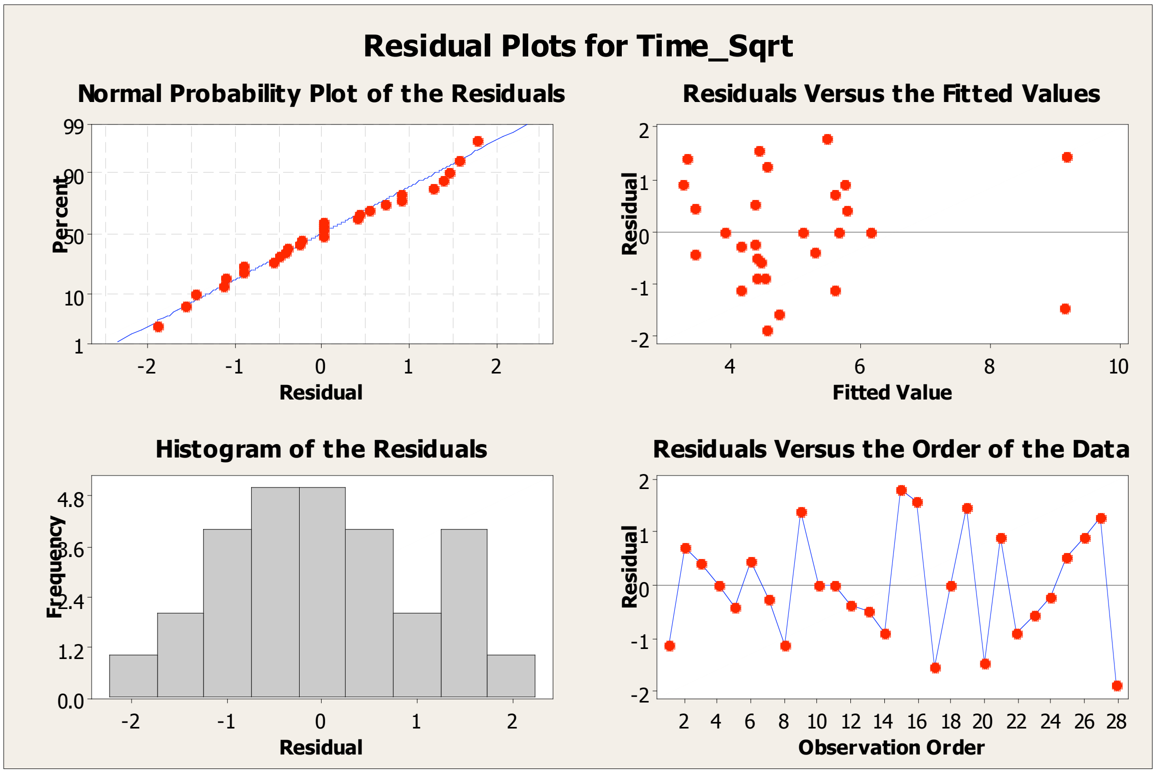


Figure 58. Residuals plots of square root of vertical error duration at MDA.

<u>Lateral errors</u>

A data summary for lateral error is shown in Figure 59. The data shows a skew to the left and high kurtosis. A normal score transformation succeeded in normalizing the data, although this will make the actual values of the transformed variable less meaningful.

AN ANOVA was run on the lateral error normal scores. The results are shown in Table 28. All variables were significant to the _=0.05 level. A check of interactions (Figure 60) shows an interaction between location and entry. As a result, this interaction was included in the model. When the model was run again, the means across the levels of the entry variable could no longer be distinguished. Figure 61 shows that the residuals were normally distributed and no departures from constant variance were noted.

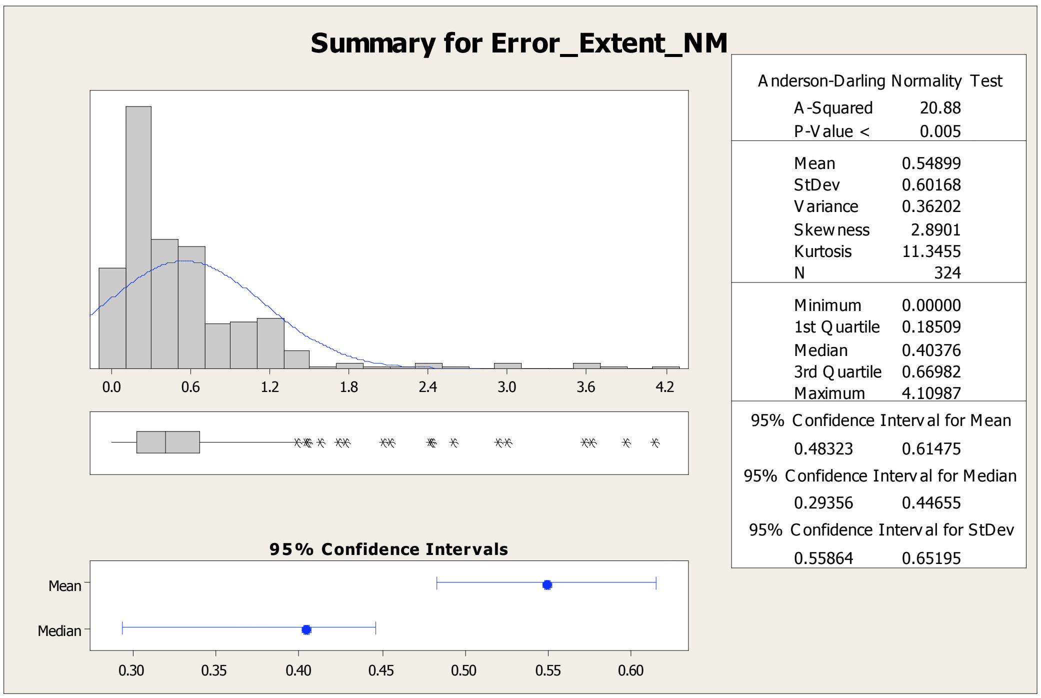


Figure 59. Data summary for lateral error.

Table 28. ANOVA for lateral error normal score.

```
Analysis of Variance for Error_NormScore, using Adjusted SS for Tests

Source                  DF    Seq SS    Adj SS    Adj MS      F      P
Subject                 14   42.1582   42.7764    3.0555   4.47  0.000
Display                  1    3.4733    3.3063    3.3063   4.83  0.029
Order                    5   10.1737   10.5152    2.1030   3.07  0.010
EntryNumber              2    4.5195    2.8653    1.4327   2.09  0.125
Location                 2   34.2315   32.7398   16.3699  23.92  0.000
EntryNumber*Location     4   18.4494   18.4494    4.6123   6.74  0.000
Error                  295  201.8667  201.8667    0.6843
Total                  323  314.8723
```

Comparisons between levels of subject showed that subject 3 was significantly different than all other subjects ($p<0.0065$) except for subject 11. All levels of location of error were significant ($p<0.04$), and there were significant differences detected between the first and fifth runs ($p=0.0255$) and the fourth and fifth runs ($p=0.0082$). Pairwise comparisons between the interaction term yielded the results shown in Table 29. Significant results are in bold, and marginally significant results are shown in italics.

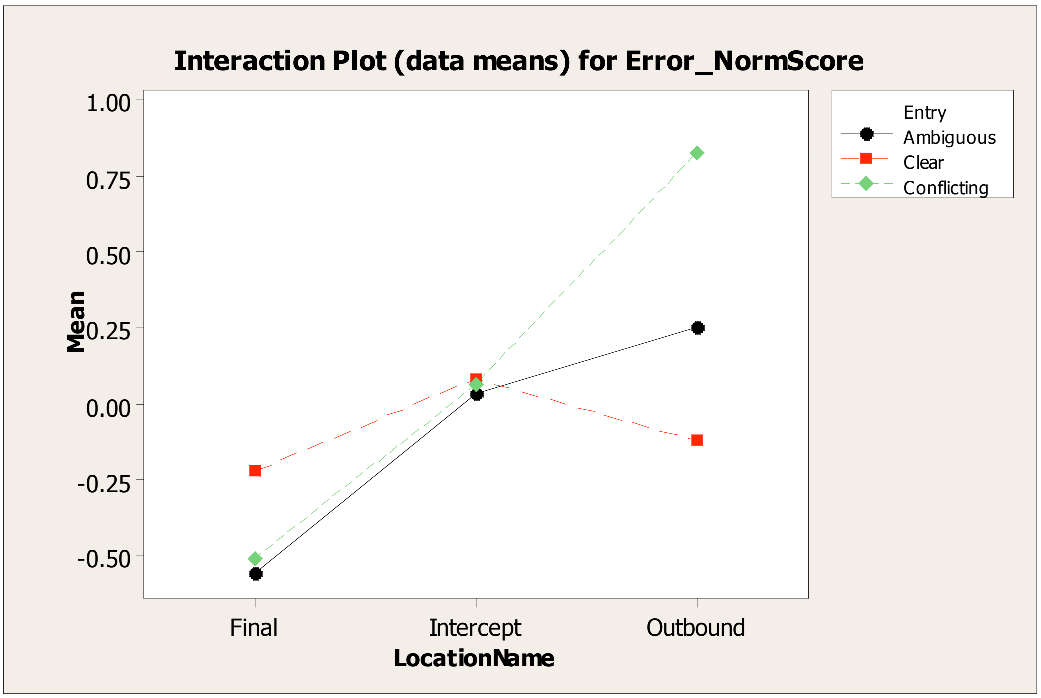


Figure 60. Interaction plot for lateral error normal score between location and entry.

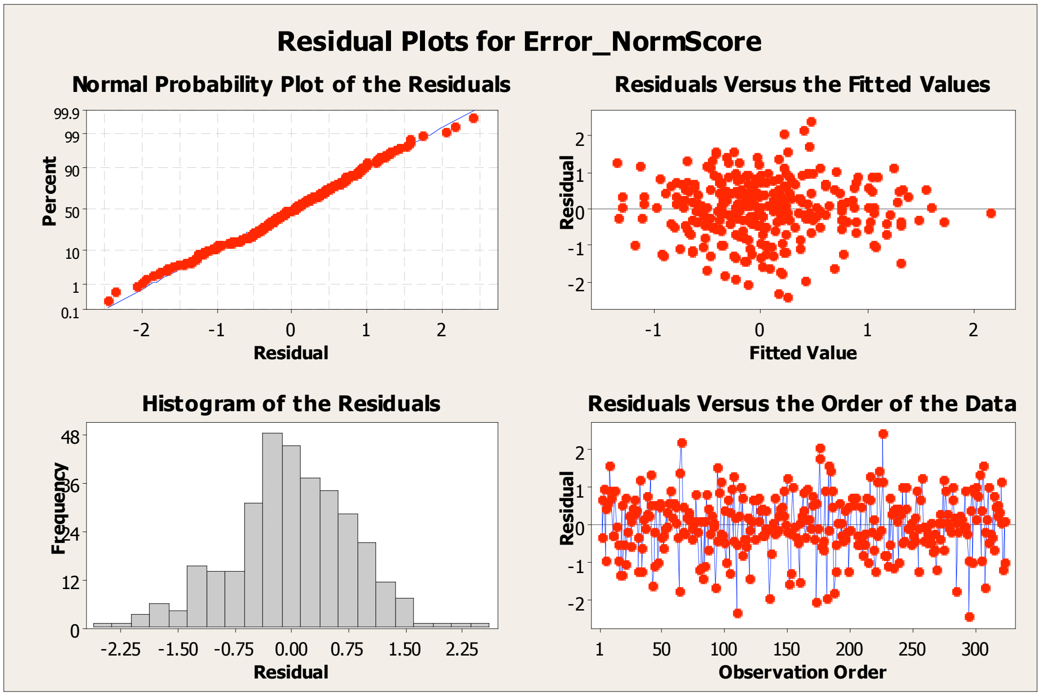


Figure 61. Residual plots for normal score of lateral error.

Table 29. Pairwise comparison p-values for lateral error interaction term.

	Clear_Outbound	Clear_Intercept	Clear_Final	Ambiguous_Outbound	Ambiguous_Intercept	Ambiguous_Final	Conflicting_Outbound	Conflicting_Intercept	Conflicting_Final
Clear_Outbound	NA	0.9710	1.0000	0.5200	0.9972	0.4207	**0.0000**	0.9908	0.4286
Clear_Intercept		NA	0.8919	0.9954	1.0000	**0.0297**	**0.0028**	1.0000	**0.0317**
Clear_Final			NA	0.3129	0.9763	0.6148	**0.0000**	0.9500	0.6230
Ambiguous_Outbound				NA	0.9567	**0.0006**	**0.0329**	0.9844	**0.0007**
Ambiguous_Intercept					NA	*0.0768*	**0.0006**	1.0000	*0.0804*
Ambiguous_Final						NA	**0.0000**	*0.0540*	1.0000
Conflicting_Outbound							NA	**0.0015**	**0.0000**
Conflicting_Intercept								NA	*0.0579*

The main effects plot shown in Figure 62 shows that greater error occurred during later portions of the approach and when using the baseline display.

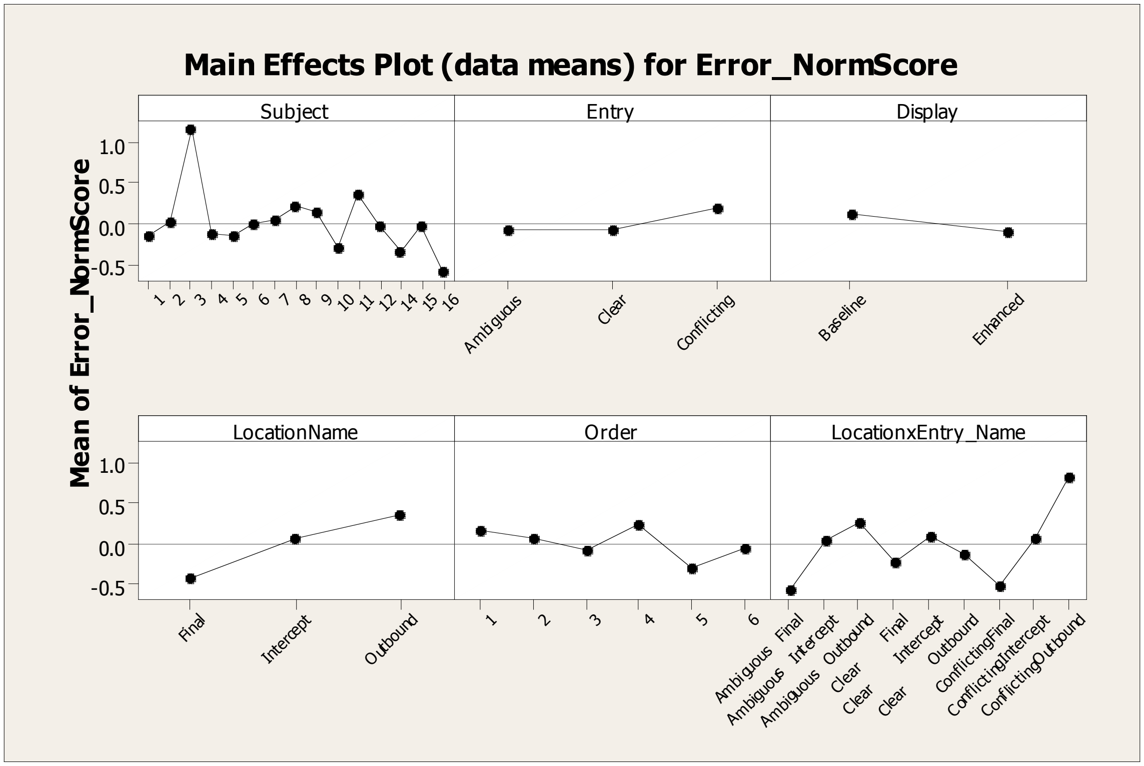


Figure 62. Main effects plot for lateral error normal score.

Off-nominal situations

As mentioned previously, subjects 7 through 16 were presented with a seventh scenario in which an engine failure forced them to abandon or deviate from the procedure. For the latter choice, these cases would represent cases of intentional noncompliance, as opposed to the unintentional noncompliance addressed by the analysis of error above. In addition to these engine-failure cases were several cases from the nominal approaches in which pilots deviated intentionally from the written procedure. This section will discuss the results of those cases.

Intentional noncompliance during non-engine failure scenarios

When given the conflicting entry scenario (for both display levels), three subjects did not turn outbound at the IAF. Outbound in those scenarios would be a left turn.

Instead, they turned right nearly 270º, then proceeded outbound and commenced the approach. The reason stated in every instance (6 in all – 3 subjects, 2 scenarios each) was to minimize the time on the "non-maneuvering" or "non-protected" side.

In addition, one subject performed a "reverse racetrack". For the conflicting entry scenario this subject turned left, but did not intercept the course outbound. Instead, the participant paralleled the course outbound, and then turned left to intercept the inbound course to the FAF. This is in violation of the procedure, since a right turn is required back inbound. However, a right turn from the position this participant encountered would have put the aircraft outside of or dangerously close to the boundary of the protected airspace for the approach.

Intentional noncompliance during engine failure scenarios

When given the engine failure, all subjects chose to turn and attempt to intercept the final approach course off of the NDB. 5 of the subjects intercepted the final course, but only 2 maintained it for any period of time, and those subjects eventually drifted off course. No subjects made it to the runway.

Recall score for intentional noncompliance scenarios

The above scenarios were added to the engine failure scenarios, and recall of lateral position was scored in an identical manner as for the other scenarios. The dotplot of the scores for the nine deviant scenarios are shown in Figure 63. The data passed tests of normality despite the small number of data points. The ANOVA is shown in Table 30, and indicates that the difference in means between the levels of the entry variable were marginally significant. Figure 64 indicates an interaction between display and entry, so this interaction was included in the model.

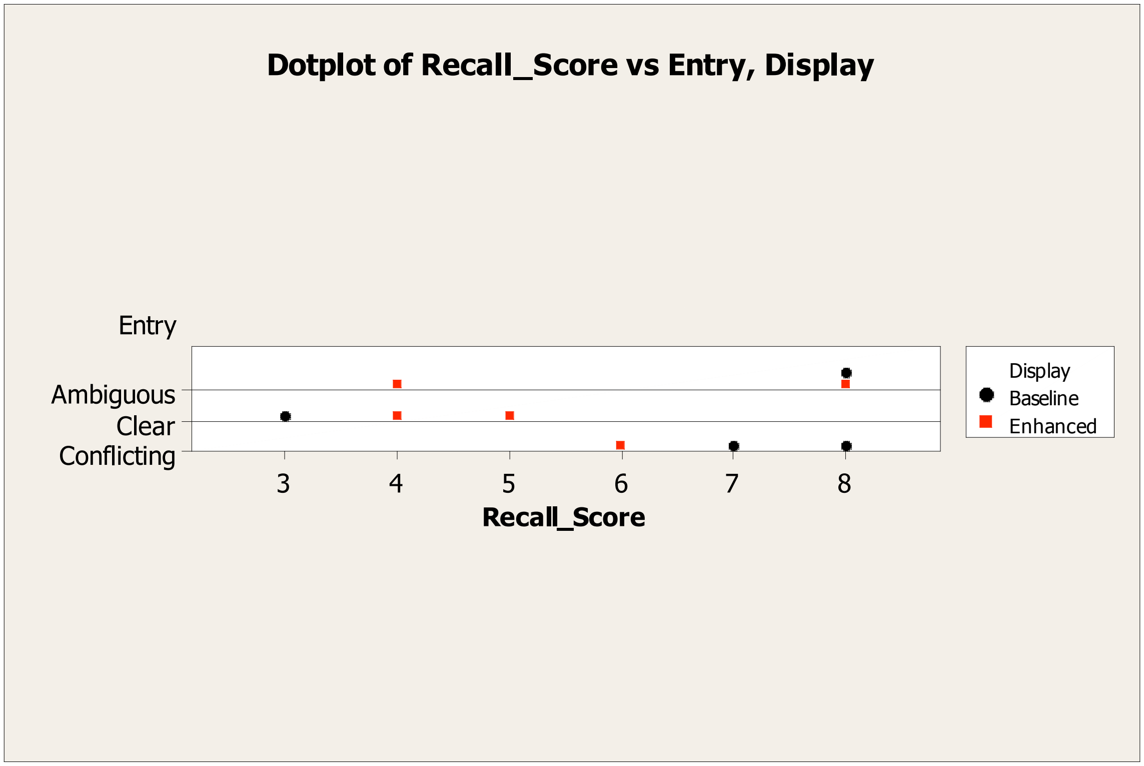


Figure 63. Dotplot of recall scores by entry and display.

Table 30. ANOVA of recall score for intentional noncompliance scenarios

```
Analysis of Variance for Recall_Score, using Adjusted SS for Tests

Source          DF   Seq SS   Adj SS   Adj MS     F      P
Display          1    3.667    0.729    0.729  0.32  0.583
Entry            2   12.383   16.143    8.072  3.55  0.065
Display*Entry    2    4.681    4.681    2.340  1.03  0.390
Error           11   25.033   25.033    2.276
Total           16   45.765
```

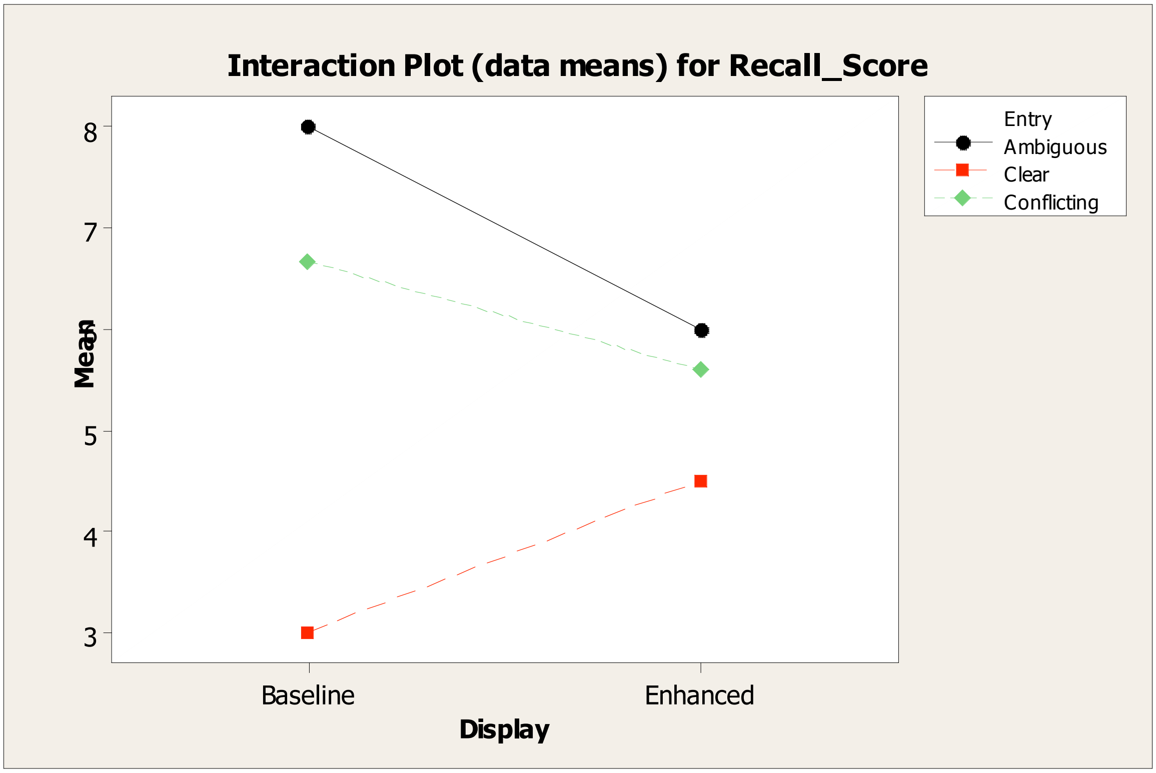


Figure 64. Interaction plot of display and entry for intentional noncompliance recall score.

The plot of residuals (Figure 65) shows no departures from normality or constant variance. Figure 66 shows the main effects for the entry variable, and pairwise comparisons indicated that the means for the clear entry were marginally different ($p<0.10$) than for the ambiguous or conflicting entries.

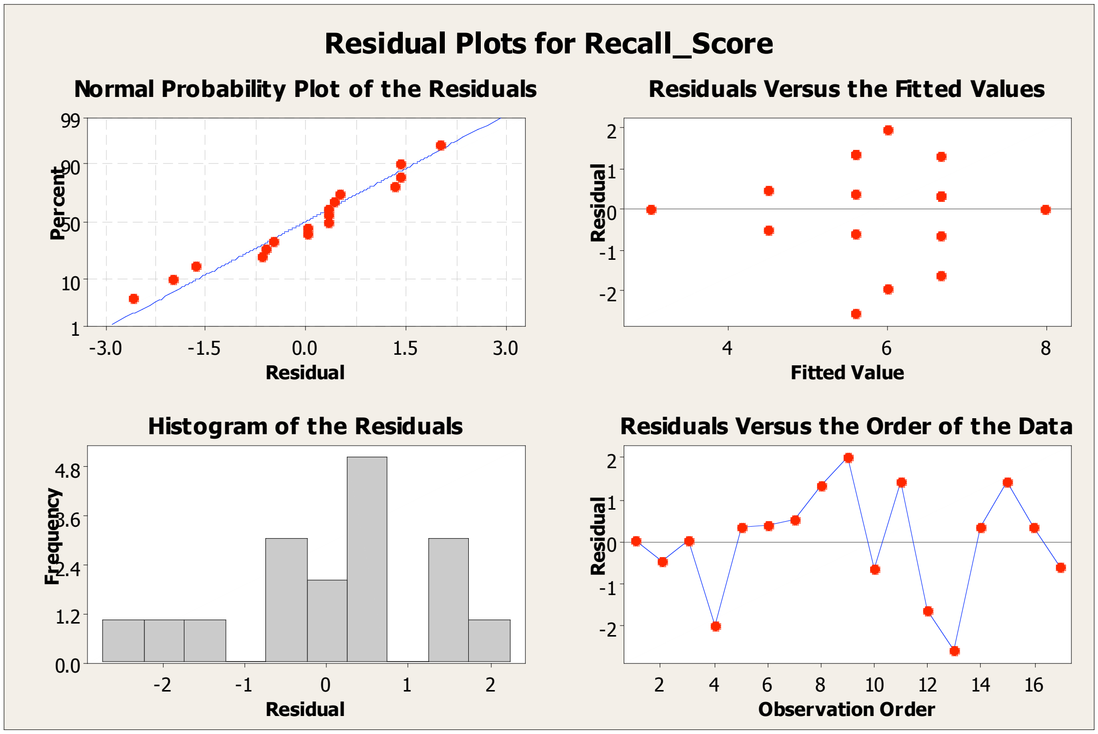


Figure 65. Residual plots for intentional noncompliance recall score ANOVA.

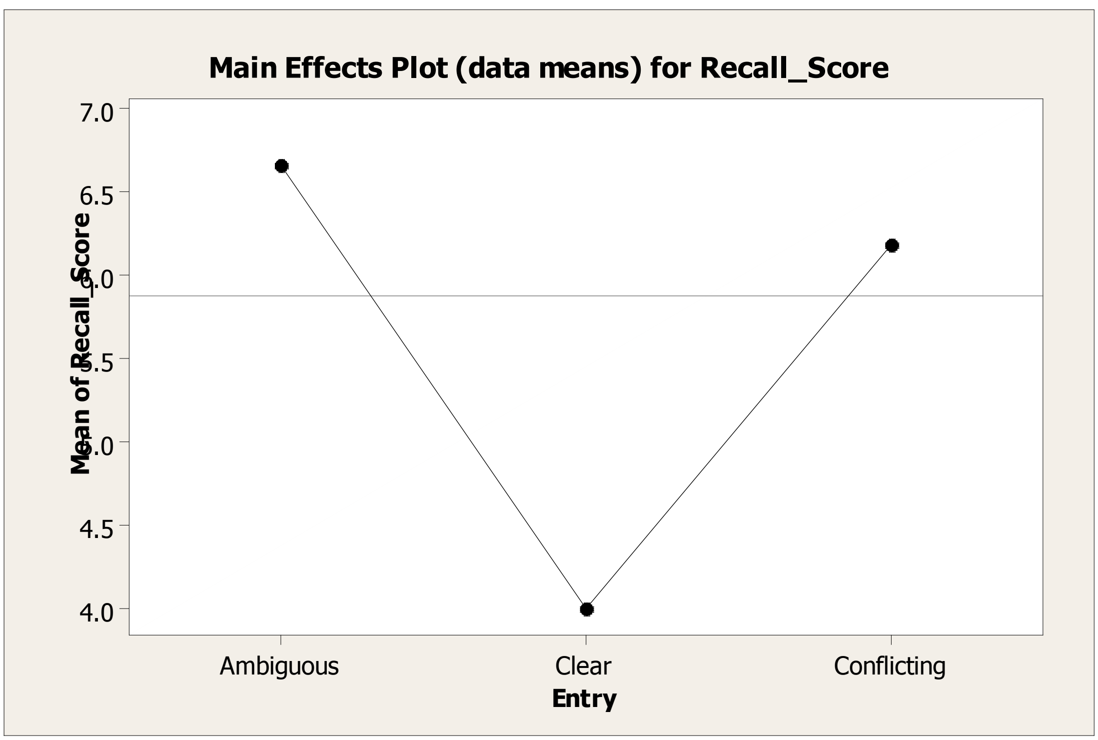


Figure 66. Means for intentional noncompliance recall score by entry.

Workload

Workload was rated using the NASA TLX scale as previously. The results were again normalized across subject. The boxplot of workload scores by display are shown in Figure 67. The boxplot of workload scores by entry are shown in Figure 68.

The boxplots do not appear to indicate any identifiable effect by display or effort. Due to the small number of data points (one per subject for subjects 7 through 16), nonparametric tests were used, but did not detect any significant differences by entry or display.

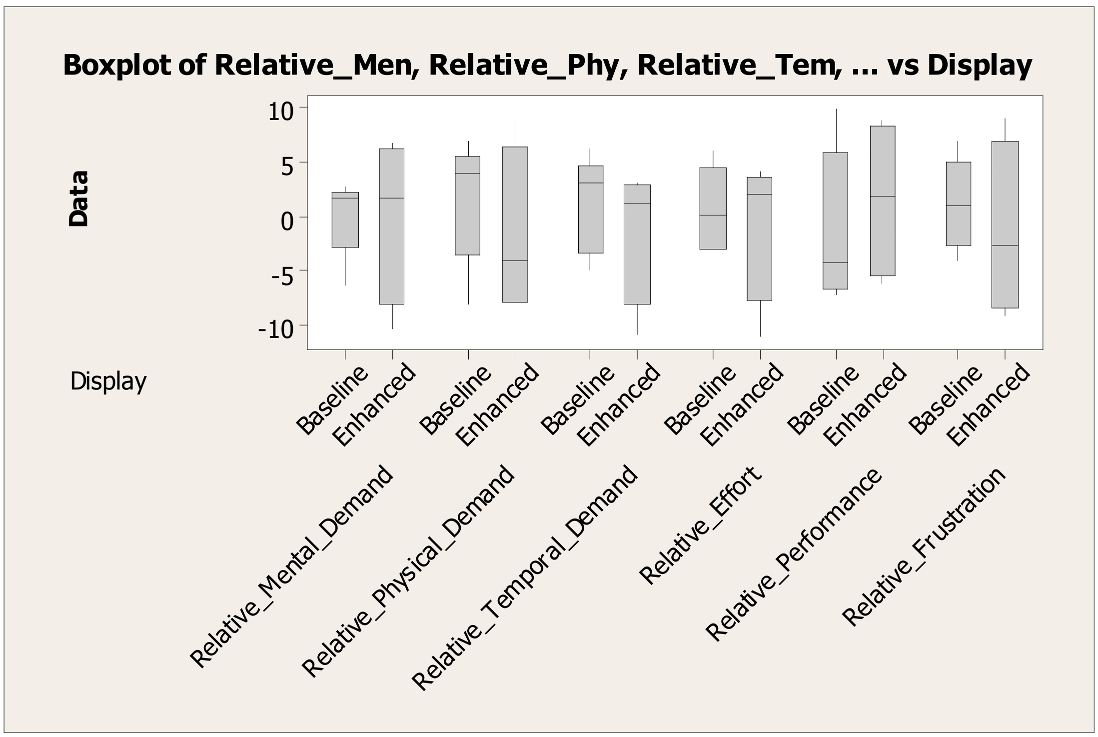


Figure 67. Boxplot of workload scores by display level.

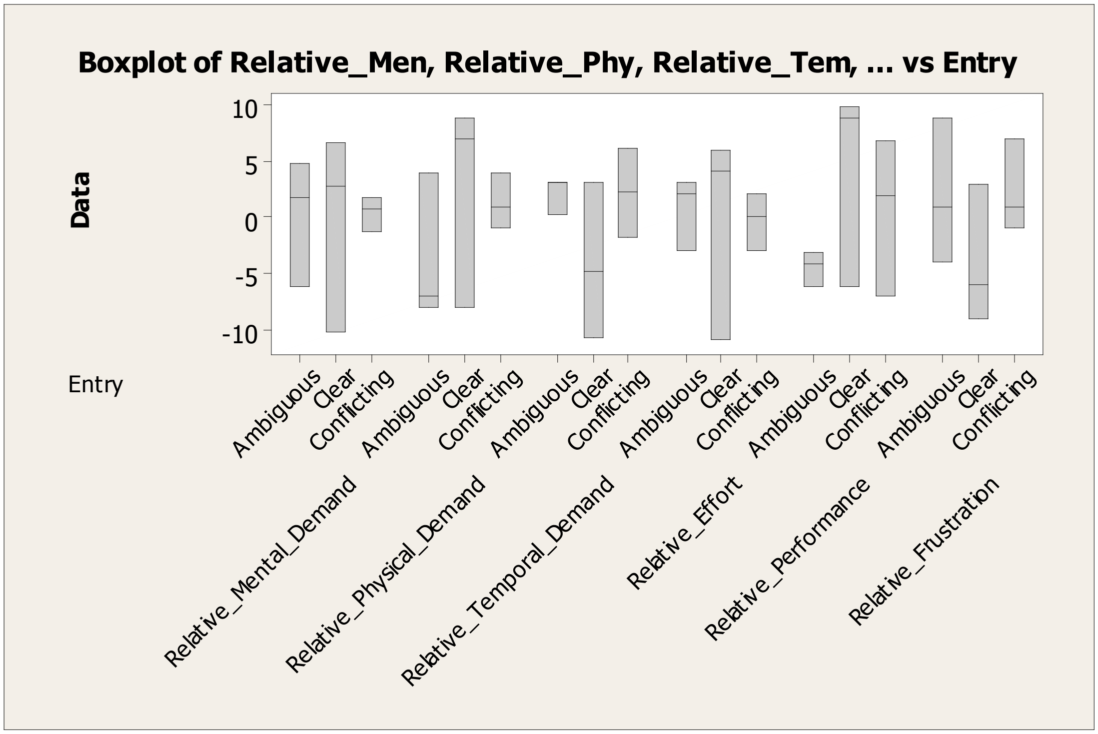


Figure 68. Boxplot of workload scores by entry.

Summary

Table 31 summarizes the main results from this section. In addition, the following findings will be discussed in the next chapter:

- All large errors on the non-maneuvering side occurred without the display of procedure context.
- Maneuvers other than the 45-180 were used in only 6 out of 90 runs.
- Three subjects used maneuvers that intentionally violated the procedure; these occurred for both display cases with the conflicting entry.
- When given an engine failure, no subjects completely abandoned the procedure; instead, subjects tried to track the inbound course.

- Statistical tests failed to detect an improvement for the enhanced display on the engine failure scenarios.

Table 31. Summary of results.

Measure	Nominal Scenario Result	Off-Nominal Scenario Result	Hypothesis supported (nominal/off-nominal)?
Recall of lateral position	Enhanced display improved recall (p=0.020)	No effect	Yes/No
Recall of vertical position	Enhanced display improved recall (p<0.001)	No effect	Yes/No
Recall of speed deviations	No effect	No effect	No/No
Workload	No effect	No effect	Yes/Yes
Control movements	No effect	No effect	Yes/Yes
Altitude error	No effect	No effect	No/Yes
Lateral error	Less error for enhanced display (p=0.029)	No effect	Yes/No

Chapter 8

Discussion of Results

The results support several important and novel conclusions. First, the results show that a display enhanced with procedure context information improved subjects' situation awareness and performance during nominal operation of the procedure turn IAP. This suggests that adding procedure context can reduce potentially dangerous procedure-following errors that have been shown to be prevalent in aviation and other domains.

Secondly, the results demonstrate that dynamic displays can be used for this purpose. Pilots were able to understand and utilize the information on the displays without additional workload. Due to experimental concerns, the additional elements represented modest (and perhaps not optimal) changes to the static display, but were still able to produce a measurable improvement. Applying a more principled analysis would likely yield even better results.

In addition, the results demonstrate that not only will pilots not comply with procedures in ways that are difficult to predict, but also that noncompliance may be a normal and safe adaptation to unusual task conditions. The subject who performed the "wrong-way" racetrack, although violating the procedure, actually made the correct choice. The subject's aircraft was on the non-protected side of the approach course, and a procedurally correct right turn would have extended the aircraft further into the non-protected airspace and perhaps outside of it (as was nearly done in other cases). So the noncompliant left turn was safer than the compliant right turn. This is in contradiction to the normal categorization of noncompliance as always wrong.

The display did not appear to change pilot's behavior with regard to noncompliance, however. In the cases of noncompliance, and in the cases of ill-advised compliance to procedure, the actions of the pilots were the same regardless of display. It was hoped that the display would have some effect on behavior, but the few instances of noncompliance, and the modest changes made to the displays (necessary to preserve controllability of the experiment), probably prevented any effects from presenting themselves.

Another interesting finding is that a pilot's ability to comprehend noncompliance appears to be limited. While interested in and able to detect noncompliance, she or he may not be able to interpret the consequences of that noncompliance. This suggests that the design of displays and procedures should consider providing support for comprehension of the implications of noncompliance, and not just for detecting noncompliance.

In addition, the results demonstrate that pilots attempted to use procedure information even when clearly outside the scope of the procedure. In the voluntary cases of noncompliance and in the impelled cases (engine failure scenarios), pilots still attempted to follow the relevant parts of the procedure. In some cases, noncompliance may have been a necessary (or at least acceptable) adaptation to the specific task conditions, which contradicts the typical assumption that all noncompliance is undesirable. This means that procedures and procedure-support aids should consider operation outside of its normal bounds in their design, rather than only for nominal operation as is currently the case.

The results failed to provide evidence that the enhanced display supported situation awareness or performance in such off-nominal situations. Although the display did not increase workload during off-nominal situations, statistical tests failed to conclude that the improvements in situation awareness and safety caused by the display were significant. This is likely due to the small number of cases induced by the experiment design, and the large variance that typifies the aviation domain.

Finally, the results extend the findings of Ockerman (2000), where a static display of procedure context was found to reduce over-reliance in an inspection task. Here the task was dynamic, and procedure context was able to reduce error and increase situation awareness.

General results

The results of the dissertation research confirm the results of the preliminary experiment in that subjects used a heuristic strategy to fly the approach. Only four subjects varied their method of accomplishing the procedure, despite the entries being specifically chosen for their convenient alignment with different maneuvers. Most subjects chose one maneuver with which they were familiar, and used that regardless of its convenience.

The use of the 45-180 maneuver may be due to its being explicitly depicted on the approach plate, which was reviewed carefully and mentioned by every subject when practicing and preparing for the approaches. In addition, the 45-180 maneuver is the most common entry used in training, so pilots may be most familiar with that entry. However, since training may vary, this finding may not hold if the experiment were repeated with a different pool of pilots, as this experiment was limited to one flying club's members.

The preliminary experiment also found that pilots were unable to comprehend the consequences of noncompliance for the procedure. In the dissertation research, three pilots utilized non-standard maneuvers for entering the procedure and for course reversal. These maneuvers were accomplished for the conflicting entry only, but for both display cases that used that entry.

One of the pilots made a "reverse" racetrack, where the outbound leg was on the unprotected side of the approach course. This is not necessarily in violation of the procedure. However, the turn inbound is depicted on the approach plate; it should be a right turn, so the left hand turn made by this pilot is in violation of the procedure. Despite this, the decision by the pilot was safer than a right turn. A different pilot who flew outbound on the unprotected side but made a right turn came within _ mile of the primary obstacle clearance area. In this case the decision to violate procedure may have been a good one.

The two other pilots who performed non-standard entries did not initially turn outbound; instead, these pilots made a 270º turn to align themselves better for the 45-180 maneuver outbound. One of these pilots explicitly mentioned that this was to prevent straying too far on the unprotected side. However, the maneuver would result in the pilots crossing the IAF twice, the second time after a delay while the 270º turn is completed. The air traffic controller (who may not have an indication on radar of where the airplanes on approach are) would have expected that this airspace to be vacated by this time. The delay may place aircraft following on the approach too close to the aircraft ahead. In this case the noncompliance, although well-intentioned, was not advisable. These pilots, much like pilots in the preliminary experiment, did not fully comprehend the consequences of their noncompliance.

It had been intended that the conflicting entry scenario would produce frequent intentional deviations from procedure instead of only for these three pilots. However, the subjects' adherence to the one-strategy approach appears to have prevented this. This strategy, while perhaps not optimal (it lead to the one aircraft coming within _ mile of the edge of the primary obstacle clearance airspace), has two benefits: (1) it can be practiced repeatedly to gain skill at implementing it, and (2) subjects do not have to try to interpret the effects of a new strategy.

Enhanced display improved situation awareness for nominal situations

The enhanced display improved subjects' recall of altitude and lateral position for the nominal situations, suggesting that subjects' situation awareness was improved by the enhanced display. By improving situation awareness, it is hoped that pilots would be better able to correct for errors. The finding that lateral errors were reduced further supports this assertion.

For the off-nominal situations, the enhanced display did not appear to improve recall. The small number of off-nominal cases (the 9 engine-failure scenarios plus the 4 cases of intentional noncompliance during the nominal scenarios) made finding results difficult. However, little indication was given that a larger number of trials would have generated a significant result.

The results do not provide any suggestions as to why this would be the case. In off-nominal situations, subjects rely on their knowledge of the system to interpret the effect of their actions (Rasmussen, 1986). Although information to support this behavior was provided (e.g. the indications of where the protected airspace was), either the information was not sufficient, or the pilot did not have enough time to consider it adequately. If the former were the case, additional information (such as more clearly interpreted navigation information) would help; for the latter case, a reduction in workload would help.

The enhanced display did not affect workload

The presence of the enhanced display did not increase either subjective workload assessments or objective indications of workload (as given by control movements) in any case. The subjective measures of workload, however, did show indications of order effect in that earlier experimental runs were rated as more difficult and as incurring lower performance. However, no indication of order effect was seen in any of the objective measures such as error or control movements. This suggests that the subjective workload impression of the subjects did not affect their performance.

Better displays of procedure context would be expected to reduce workload. For the purposes of this study, however, a reduction in workload would have, by itself, improved performance and recall. This would cause the effect of reduced workload to be confounded with the effect of the display elements. The finding that workload was not altered by the display therefore supports the notion that the results were due to the presence of the information and not any particular effect on the task itself.

The subjective workload measures were overall measures (not broken down by position on the approach) and varied by subject. The objective measures of workload were given by the control movements, with greater variance in control movements indicative of higher workload. These also varied by subject, but could be broken down by position on the approach. These measures indicated a substantial increase in workload at the end of the approach which coincided with a decrease in error. This increase in workload toward the end of the approach may have caused there to be no concrete effect of the display on situation awareness of speed and missed approach errors.

The enhanced display improved safety for the nominal cases

Lateral error was reduced in the enhanced display cases for the nominal scenarios. This finding is supported by there being no significant errors on the non-maneuvering side when using the enhanced display. These results suggest that the enhanced display was successful in increasing safety, at least in the lateral plane.

As mentioned, the increase in situation awareness may explain the reduction in error, although a similar increase in situation awareness of altitude errors did not result in a reduction in those errors. Moreover, anytime there is an increase in situation awareness, it is difficult (if not impossible) to de-confound the effect on safety. In practical terms, it may not be particularly important, since the desire to increase situation awareness has as its goal to increase safety.

That a lateral effect but not an altitude effect would be found is somewhat surprising since the enhanced display seems to provide more direct support for altitude compliance rather than track compliance. No concrete evidence indicates why the display did not support altitude compliance, but three possible reasons are presented here. First, it is important to note that the scale of altitude deviations (as compared to lateral deviations) is different. Altitude deviations are measured in 100s of feet, whereas lateral deviations are typically 1000s of feet or even miles.

Secondly, there is an instrument that directly indicates altitude (the altimeter). Therefore, errors are directly displayed to the pilot, whereas lateral errors must be gleaned from cross-checking two instruments (at least for this task). This makes altitude errors much more obvious than lateral errors, with or without procedure context elements.

In addition, the simulator's characteristics were different for the two axes. The simulated aircraft was very stable laterally, but not very stable in the pitch axis. This meant that corrections to lateral errors could be made and not tracked, whereas corrections to vertical errors needed to be continuously monitored. Since detection of deviations was not closely measured, it is not possible to prove this explanation. However, this type of control behavior could overwhelm any visible effect of the display on corrections to altitude deviations.

Implications of the findings

The findings suggest that support for procedure following in both nominal and off-nominal situations is needed. Pilots relied upon a heuristic strategy in most cases, strategies which may not be adequate in all circumstances (even if they are allowed by the procedure). In cases where pilots did not rely on these heuristic strategies, they appeared to be unable to comprehend the consequences of their noncompliance.

Displays of even simple elements of procedure context appear to help. In the experiments in this dissertation, they reduced error and increased situation awareness in nominal situations. Additional work needs to be done to determine how to extend these benefits to the off-nominal situations.

The findings should generalize to other highly proceduralized domains. As mentioned in Chapter 1, procedures and checklists are one of the most frequent contributors to accidents. The need for improvements in procedure-following is therefore similar to that of aviation. In addition, researchers in numerous domains have found that humans use heuristic strategies to accomplish tasks. It is therefore unsurprising to find it here in a high-workload task, and it would be equally unsurprising to find it in other tasks in other domains. Is such cases, procedure context information can potentially increase situation awareness and reduce error.

Chapter 9

Conclusions and Recommendations for Future Work

Aviation is one of many highly proceduralized, safety-critical domains. In these domains, procedures are relied upon for safe and efficient operations, and failure to follow procedures is frequently cited as a cause of accidents. Yet support for procedure following has been minimal and in general has not utilized the full capabilities of current technology. Before those capabilities can be leveraged, however, a disciplined approach to how technology can be used to support procedure following should be developed. This dissertation represents an early effort at defining an approach.

Procedures couple operators to their system. It defines a sequence of interactions with the system which will achieve a particular goal. As such a procedure provides a method of accomplishing the task which is known to be successful, but also provides predictability and standardization.

The predictability that procedures provide is also a source of information which is often extremely useful to other operators in the system. The preliminary experiment in this dissertation attempted to utilize that information directly, with mixed results. The pilots in that task were apparently able to comprehend the meaning of the information, but were unable to comprehend the consequences for the system when the underlying rationale for the procedure was no longer valid.

Of course, the predictability of the procedure is not absolute. In addition to the normal sources of uncertainty present in the environment, the coupling provided by procedures is dependent upon the less-than-deterministic human operator.

The operators in the experiments described in this dissertation were following procedures. In one case the procedure was novel and was designed in part to expect adherence to procedure. In a second case the procedure was well-established and well-known by the pilots. In neither case did pilots perfectly comply with procedures. They improvised based on their understanding of the situation; they did not expect that the procedure had been designed in consideration of the situation they encountered. Despite the controlled circumstances of the experimental environment, in some cases the pilots were justified in not complying with procedure.

Whether justified or not, and whether due to environmental uncertainty, their own actions, or the actions of others in the system, operators may find themselves outside of the situations envisioned by the procedure designer. In any case, this research has suggested that operators may still follow a procedure or make use of portions of a procedure. Therefore, displays for procedure following should support operation in both nominal and off-nominal situations. In systems with multiple operators, this support is even more critical.

In the dissertation research, the context of procedures, information which is utilized to design the procedures but is often not present in its implementation, was used to design an aid for pilots to fly an instrument approach. This information provided significant benefits to the pilot, increasing situation awareness and decreasing lateral navigation errors during nominal situations. However, the results were unable to demonstrate these benefits for off-nominal situations.

Future work

The results of the experiment do not suggest an explanation for why procedure context did not assist pilots during off-nominal situations. It could be that these off-nominal situations require different information than is contained in the procedure context. It could also be that the modest changes made for this experiment were insufficient to provide detectable benefits.

Future research should investigate whether procedure context is useful for off-nominal situations. Different tasks may allow more significant changes to be made to the interface, or other parts of procedure context could be added. In addition, some analysis of the utility of certain information for off-nominal situations could identify what would be helpful.

In addition, different procedure context elements, or modifications to the procedure context elements, should be investigated, particularly for dynamic displays. Hierarchical techniques used for designing displays, such as Ecological Interface Design (Vicente & Rasmussen, 1992) could be useful for this purpose.

For the two experiments described herein, no changes were made to the actual procedure. The analysis which identified procedure context elements could also identify desirable changes to the underlying procedure. The full utility of procedure context can be examined by testing such changes.

Within aviation, research should be extended to general aviation and maintenance. From 1988 to 1997, the accident rate for general aviation was over 20 times higher than for commercial operations, with over 5 times as many fatalities (Goldman, Fiedler, & King, 2002). Moreover, due to the expense of certifying and purchasing equipment, general aviation aircraft generally do not benefit from new displays. As a result, improvements in the design of flight deck procedures may be a promising method for increasing safety.

A report by Britain's Civil Aviation Authority identified aircraft maintenance as a contributing factor for 12 percent of air transport accidents from 1959 to 1983, a figure that increased in the late 1980s and early 1990s (Safety Regulation Group, 2002). Aircraft maintenance is another highly proceduralized domain and could significantly benefit from improvements in procedure support.

As mentioned, this work has implications for other highly proceduralized domains. Research should be undertaken to see how procedure context can benefit domains such as medicine, maritime, industrial, and chemical.

Appendix 1

Experiment Forms and Questionnaires

Demographic questionnaire

Male   Female

Age________

Total Hours____

Instrument Hours____

Current Aircraft____

Hours in Current Aircraft ____

Other Hours ____

License/rating____

Base Airport____

To your knowledge, have you ever operated an aircraft into or out of Millville Municipal airport in New Jersey? ____________________________

In the space below, please describe your understanding of how to fly a procedure turn.

Situation awareness questionnaire

1. On the map attached, indicate the initial position of the aircraft and the track of the aircraft, including missed approach.

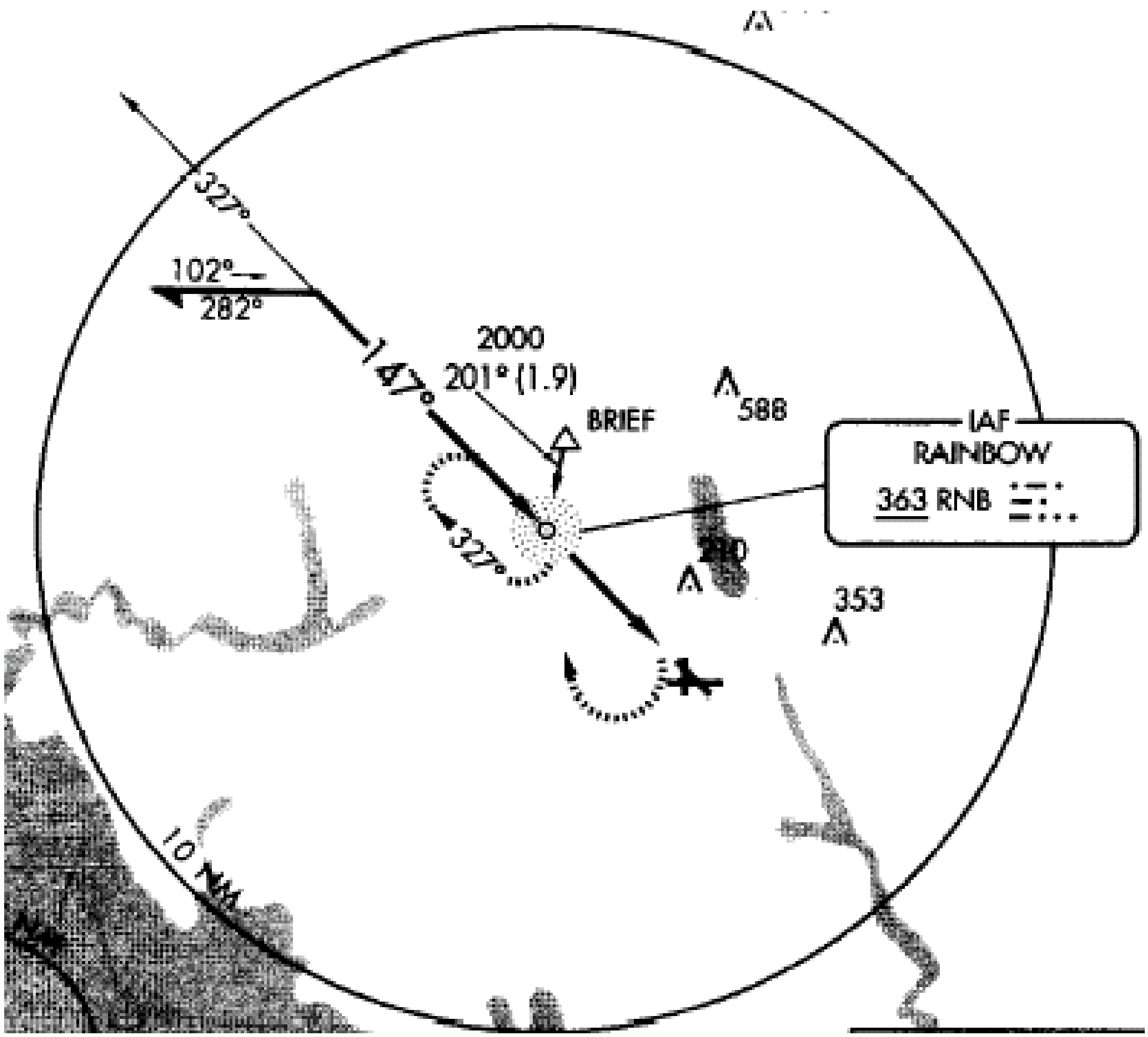
2. Did you remain above 2000 feet until established on the inbound course? If no, please indicate why you did not.

3. Did you remain above 1300 feet until past the FAF? If no, please indicate why you did not.

4. Did you remain above the MDA at all times? If no, please indicate why you did not.

5. Please indicate whether you recall dropping below approach speed. If so, please indicate the frequency, duration, and magnitude of the deviations you recall.

6. Please indicate any deviations from the missed approach procedure that you recall, and any reasons for those deviations.


327°
102°
282°
147°
2000
201° (1.9)
BRIEF
588
IAF
RAINBOW
363 RNB
327°
353
10 NM

Workload questionnaire

RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Mental Demand

Low High

Physical Demand

Low High

Temporal Demand

Low High

Effort

Low High

Performance

Low High

Frustration

Low High

Appendix 2

Additional Experiment Results

Table A2-1. Tabulated statistics of subject, entry, and display for recall of lateral trajectory.

Rows: Subject   Columns: Entry / Display

	Ambiguous		Clear		Conflicting		All
	Baseline	Enhanced	Baseline	Enhanced	Baseline	Enhanced	All
1	6.000	7.000	3.000	8.000	3.000	6.000	5.500
	6.000	7.000	3.000	8.000	3.000	6.000	6.000
	*	*	*	*	*	*	2.074
2	4.000	7.000	7.000	4.000	2.000	6.000	5.000
	4.000	7.000	7.000	4.000	2.000	6.000	5.000
	*	*	*	*	*	*	2.000
3	4.000	6.000	3.000	4.000	8.000	8.000	5.500
	4.000	6.000	3.000	4.000	8.000	8.000	5.000
	*	*	*	*	*	*	2.168
4	4.000	7.000	5.000	2.000	5.000	8.000	5.167
	4.000	7.000	5.000	2.000	5.000	8.000	5.000
	*	*	*	*	*	*	2.137
5	2.000	6.000	3.000	5.000	6.000	8.000	5.000
	2.000	6.000	3.000	5.000	6.000	8.000	5.500
	*	*	*	*	*	*	2.191
6	4.000	8.000	4.000	7.000	6.000	7.000	6.000
	4.000	8.000	4.000	7.000	6.000	7.000	6.500
	*	*	*	*	*	*	1.673
7	6.000	6.000	4.000	2.000	6.000	7.000	5.167
	6.000	6.000	4.000	2.000	6.000	7.000	6.000
	*	*	*	*	*	*	1.835
8	4.000	7.000	5.000	3.000	3.000	6.000	4.667
	4.000	7.000	5.000	3.000	3.000	6.000	4.500
	*	*	*	*	*	*	1.633
9	7.000	2.000	8.000	7.000	7.000	5.000	6.000
	7.000	2.000	8.000	7.000	7.000	5.000	7.000
	*	*	*	*	*	*	2.191
10	8.000	8.000	8.000	7.000	7.000	7.000	7.500
	8.000	8.000	8.000	7.000	7.000	7.000	7.500
	*	*	*	*	*	*	0.548
11	5.000	7.000	8.000	6.000	5.000	6.000	6.167
	5.000	7.000	8.000	6.000	5.000	6.000	6.000
	*	*	*	*	*	*	1.169
12	7.000	7.000	7.000	6.000	6.000	7.000	6.667

	7.000	7.000	7.000	6.000	6.000	7.000	7.000
	*	*	*	*	*	*	0.516
14	6.000	8.000	6.000	8.000	6.000	5.000	6.500
	6.000	8.000	6.000	8.000	6.000	5.000	6.000
	*	*	*	*	*	*	1.225
15	4.000	7.000	5.000	6.000	5.000	3.000	5.000
	4.000	7.000	5.000	6.000	5.000	3.000	5.000
	*	*	*	*	*	*	1.414
16	8.000	8.000	5.000	6.000	5.000	6.000	6.333
	8.000	8.000	5.000	6.000	5.000	6.000	6.000
	*	*	*	*	*	*	1.366
All	5.267	6.733	5.400	5.400	5.333	6.333	5.744
	5.000	7.000	5.000	6.000	6.000	6.000	6.000
	1.751	1.486	1.844	1.993	1.633	1.345	1.739

Cell Contents: Recall_Score : Mean
Recall_Score : Median
Recall_Score : Standard deviation

Table A2-2. Tabulated statistics for subjects, entry, and display for altitude recall.

	Ambiguous		Clear		Conflicting		All
	Baseline	Enhanced	Baseline	Enhanced	Baseline	Enhanced	All
1	3.000	3.000	2.000	3.000	2.000	3.000	2.667
	3.000	3.000	2.000	3.000	2.000	3.000	3.000
	*	*	*	*	*	*	0.5164
	1	1	1	1	1	1	6
2	3.000	3.000	3.000	2.000	2.000	3.000	2.667
	3.000	3.000	3.000	2.000	2.000	3.000	3.000
	*	*	*	*	*	*	0.5164
	1	1	1	1	1	1	6
3	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.0000
	1	1	1	1	1	1	6
4	1.000	2.000	3.000	3.000	3.000	3.000	2.500
	1.000	2.000	3.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.8367
	1	1	1	1	1	1	6
5	2.000	2.000	1.000	3.000	1.000	3.000	2.000
	2.000	2.000	1.000	3.000	1.000	3.000	2.000
	*	*	*	*	*	*	0.8944
	1	1	1	1	1	1	6
6	2.000	3.000	2.000	3.000	3.000	2.000	2.500
	2.000	3.000	2.000	3.000	3.000	2.000	2.500
	*	*	*	*	*	*	0.5477
	1	1	1	1	1	1	6
7	1.000	3.000	2.000	3.000	2.000	3.000	2.333
	1.000	3.000	2.000	3.000	2.000	3.000	2.500

	*	*	*	*	*	*	0.8165
	1	1	1	1	1	1	6
8	1.000	2.000	3.000	3.000	1.000	2.000	2.000
	1.000	2.000	3.000	3.000	1.000	2.000	2.000
	*	*	*	*	*	*	0.8944
	1	1	1	1	1	1	6
9	2.000	3.000	2.000	3.000	3.000	3.000	2.667
	2.000	3.000	2.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.5164
	1	1	1	1	1	1	6
10	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.0000
	1	1	1	1	1	1	6
11	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.0000
	1	1	1	1	1	1	6
12	3.000	2.000	3.000	3.000	2.000	3.000	2.667
	3.000	2.000	3.000	3.000	2.000	3.000	3.000
	*	*	*	*	*	*	0.5164
	1	1	1	1	1	1	6
14	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.0000
	1	1	1	1	1	1	6
15	3.000	3.000	2.000	3.000	3.000	3.000	2.833
	3.000	3.000	2.000	3.000	3.000	3.000	3.000
	*	*	*	*	*	*	0.4082
	1	1	1	1	1	1	6
16	3.000	3.000	2.000	3.000	2.000	3.000	2.667
	3.000	3.000	2.000	3.000	2.000	3.000	3.000
	*	*	*	*	*	*	0.5164
	1	1	1	1	1	1	6
All	2.400	2.733	2.467	2.933	2.400	2.867	2.633
	3.000	3.000	3.000	3.000	3.000	3.000	3.000
	0.8281	0.4577	0.6399	0.2582	0.7368	0.3519	0.6080
	15	15	15	15	15	15	90

```
Cell Contents:  Vert_Score  :  Mean
                Vert_Score  :  Median
                Vert_Score  :  Standard deviation
                               Count
```

References

Aviation Safety Network (2003). Aircraft accident description 23 AUG 2000 Gulf Air Airbus A.320 A40-EK. Retrieved 26 April 2004, from http://aviation-safety.net/database/2000/000823-0.htm.

Billings, C. E. (1995). Situation awareness measurement and analysis: A commentary. *Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness*. Daytona Beach, FL: Embry Riddle University Press, 1-6.

Bolman, L. (1979). Aviation accidents and the "Theory of the Situation." In G. E. Cooper, M. D.White, and J. K. Lauber (Eds.), Resource management in the cockpit. Moffett Field, CA: NASA Ames Research Center (NASA Conference Publication 2120).

Chandra, D.C., Mangold, S.J. & Riley, V. (2002). H*uman factors considerations in the design and evaluation of electronic flight bags (EFBs) Version 2*. Retrieved 26 April 2004, from: http://www.volpe.dot.gov/opsad/efb/vreppub.html.

Converse, S. A. (1994). Operating procedures: Do they reduce operator errors? *Proceedings of the Human Factors and Ergonomics Society 28th Annual Meeting*. Santa Monica, CA: The Human Factors Society.

Crane, P. M. (1992). Theories of expertise as models for understanding situation awareness. In *Proceedings of the 13th Annual Symposium on Psychology in the Department of Defense*, 148-142.

Degani, A., & Weiner, E. (1990). *Human factors of flight-deck checklists: The normal checklist* (NASA Contractor Report 177549), NASA: Ames Research.

Degani, A. & Wiener, E. (1994). *On the design of flight deck procedures* (NASA Contractor Report 177642). NASA: Ames Research Center.

Degani, A., & Wiener, E. (1997). Procedures in complex systems: The airline cockpit. The airline cockpit. IEEE Transactions on Systems, Man, and Cybernetics, 27(3), 302-312.

Endsley, M. R. (1999). Situation awareness in aviation systems. In Garland, D. J., Wise, J. A., and Hopkin, V. D. (Eds.), *Handbook of Aviation Human Factors,* Mahwah, NJ: Lawrence Erlbaum, 257-276.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley and D. J. Garland (Eds) *Situation Awareness Analysis and Measurement,* 2000, London: Lawrence Erlbaum, 113-128.

Federal Aviation Administration (1994a). *Federal Aviation Regulations Part 91: General Operating and Flight Rules*. Washington, DC: Federal Aviation Administration.

Federal Aviation Administration (1994b). *Federal Aviation Regulations Part 97: Standard Instrument Approach Procedures.* Washington, DC: Federal Aviation Administration.

Federal Aviation Administration (2002). *Aeronautical Information Manual: Official Guide to Basic Flight Information and ATC Procedures.* Washington, DC: Federal Aviation Administration, Section 5-4-8a.

Federal Aviation Administration (2003). *U.S. Terminal Procedures*. Washington, DC: Federal Aviation Administration, Volume NE-2, 167.

Fracker, M. L. (1991). *Measures of Situation Awareness: Review and Future Directions* (AL-TR-1991-0128). Wright-Patterson AFB, OH: Amstrong Laboratory.

Goldman, S. M., Fiedler, E. R., and King, R. E. (2002). General aviation maintenance-related accidents: A review of ten years of NTSB data. *DOT/FAA/AM-02/23*. Oklahoma City: Federal Aviation Administration.

Gross, R. L. (1995). Studies suggest methods for optimizing checklist design and crew performance. *Flight Safety Digest*, *14*(5), 1-10.

Hacker, E. W. (1984). *Learning from the Past: A Fighter Pilot's Obligation*, Retrieved 26 April 2004, from http://www.globalsecurity.org/military/library/report/1984/HEW.htm.

Hammer, J. (1999). Case study of paired approach procedure to closely spaced parallel runways. *Air Traffic Control Quarterly*, *8*(3), 223-252.

Hammer, J. M. (1984). An intelligent flight-management aid for procedure execution. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-14*(6), 885-888.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North-Holland, 139-183.

Harwood, K., Barnett, B., & Wickens, C. (1988). Situational awareness: A conceptual and methodological framework. Proceedings of the Symposium Psychology in the Department of Defense.

Haynes, A. (1991). Presented at NASA Dryden Flight Research Facility. Retrieved 26 April 2004, from http://www.panix.com/~jac/aviation/haynes.html.

Hughes, J. A. (1992). Faltering from ethnography to design. *Proceedings of the {ACM} Conference on Computer Supported Cooperative Work (CSCW)'92*, Toronto, Ontario: ACM Press.

Ippolito, C. & Pritchett, A. (2000). Software architecture for a reconfigurable flight simulator. AIAA Modeling and Simulation Technologies Conference: Denver, CO.

Jones, D. G., & Endsley, M. R. (2000). Can real-time probes provide a valid measure of situation awareness? In *Proceedings of the Human Performance, Situation Awareness and Automation: User Centered Design for the New Millennium Conference*.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. *Human Factors*, *35*(2), 221-242.

Knaeuper, A., & Morris, N. M. (1984). A model-based approach for online aiding and training in process control. *Proceedings of the 1984 IEEE Conference on Systems, Man, and Cybernetics*. Halifax, Nova Scotia: IEEE.

Landry, S. J. & Pritchett, A. R. (2000). The safe zone for paired closely spaced parallel approaches: Implications for procedures and automation. *Proceedings of the 19th Digital Avionics Systems Conference*, 1, 3E3/1-3E3/8.

Landry, S. J. & Pritchett, A. R. (2002a). Incorporating procedural information in a paired approach task. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 106-114.

Landry, S. J. & Pritchett, A. R. (2002b). Displaying procedural vs. real-time information for paired approaches. *Proceedings of the 21st Digital Avionics Systems Conference*, 2, 11B3-1 – 11B3-9.

Linde, C., & Goguen, J. (1987). *Checklist interruption and resumption: A linguistic study* (NASA contract report 177460). Moffett Field, CA: NASA Ames Research Center.

Marsden, P. Procedures in the nuclear industry. (1996). In N. Stanton (Ed.), *Human Factors in Nuclear Safety*, London: Taylor and Francis, 99-116.

McRuer, D. & Graham, D. (1965). Human pilot dynamics in compensatory systems, *AFFDL-TR-65-15*.

Monan, W. P. (1979). *Distraction – A human factor in air carrier hazard events* (NASA Technical memorandum 78608). Moffett Field, CA: NASA Ames Research Center.

National Transportation Safety Board. (1969). Pan American World Airways. Boeing B-707 N799PA. Anchorage, Alaska. December 12, 1968. Washington, DC: NTSB.

National Transportation Safety Board. (1983a). *DCA83AA014*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X41968&key=1.

National Transportation Safety Board. (1983b). *MIA83AA136*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X43070&key=1.

National Transportation Safety Board. (1987a). *DCA87MA033*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X30906&key=1.

National Transportation Safety Board. (1987b). *DCA87MA046*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X31759&key=1.

National Transportation Safety Board. (1988a). *CHI88IA233*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X26494&key=1.

National Transportation Safety Board. (1988b). *DCA88MA072*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X26528&key=1.

National Transportation Safety Board. (1988c). *LAX88IA064*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X32714&key=1.

National Transportation Safety Board. (1989). *DCA89MA074*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X29335&key=1.

National Transportation Safety Board. (1990). *LAX90IA305*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X24039&key=1.

National Transportation Safety Board. (1994a). *CHI94IA039*, Retrieved 10 May 2003, from http://www.ntsb.gov/ntsb/brief.asp?ev_id=20001214X13714&key=1.

National Transportation Safety Board. (1994b). A review of flightcrew-involved major accidents of U.S. air carriers, 1978 through 1990, *NTSB/SS-94/01*. Washington, DC: NTSB.

National Transportation Safety Board. (2001). *Aircraft Accident Report NTSB/AAR-01/02: Runway Overrun During Landing: American Airlines Flight 1420.*. Washington, DC: National Transportation Safety Board.

National Transportation Safety Board. (2004). *National Transportation Safety Board Website*, http://www.ntsb.gov

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books, 44.

Novick, D. G. (1999). Using the cognitive walkthrough for operating procedures. *interactions*, 6(3), 31-37. New York: ACM Press.

Ockerman, J. (2000). *Task Guidance and Procedure Context: Aiding Workers in Appropriate Procedure Following*. Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA.

Ockerman, J. & Pritchett A. (1998). Task guidance in aircraft inspection. In Proceedings of The Second International Symposium on Wearable Computers. Pittsburgh, PA: IEEE, 33-40.

Ockerman, J., & Pritchett, A. (2000). A review and reappraisal of task guidance: Aiding workers in procedure following. *International Journal of Cognitive Ergonomics*, *4(*3), 191-212.

Palmer, E. & Degani, A. (1991). Electronic checklists: Evaluation of two levels of automation. In R. Jensen (Ed.), *Proceedings of the Sixth Symposium on Aviation Psychology*. Colombus, OH: The Ohio State University.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 240-253.

Perrow, C. (1984). *Normal Accidents*. New York: Basic Books.

President's Commission on the Accident at Three Mile Island. (1979). *Report of the President's Commission on the Accident at Three Mile Island*. Retrieved 26 April 2004, from http://stellar-one.com/nuclear/report_to_the_president.htm.

Pritchett, A. & Landry, S. (2001). Two studies of paired approaches. Presented at The 2001 Air Traffic Management Research and Design Workshop: Sante Fe, NM.

Pritchett, A. (1999). Pilot performance at collision avoidance during closely spaced parallel approaches. Air Traffic Control Quarterly, Vol. 7(1) , 47-75.

Pritchett, A. R., & Yankosky, L. J. (2000). Pilot performance at new ATM operations: Maintaining in-trail separation and arrival sequencing. Proceedings of the AIAA Guidance, Navigation and Control Conference. Reston, VA: American Institute of Aeronautics & Astronautics.

Pritchett, A. R., Hansman, R. J, & Johnson, E. N. (1995). Use of testable responses for performance-based measurement of situation awareness. In D. J. Garland and M. R. Endsley (Eds.), *Experimental Analysis and Measurement of Situation Awareness*, 75-81.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York: Elsevier Science.

Regal, D. M., Rogers, W. H., & Boucek, G. P. (1989). Situational awareness in the commercial flight deck – definition, measurement, and enhancement. *Proceedings of the 7th SAE Aerospace Behavioral Technology Conference*, 65-69.

Rouse, S. H. & Rouse, W. B. (1980). Computer-based manuals for procedural automation. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-10*(8).

Rouse, S. H., Rouse, W. B., & Hammer, J. M. (1982). Design and evaluation of an on-board computer based information system for aircraft. *IEEE Transactions on System, Man, and Cybernetics*, *12*(4), 451-463.

RTCA SC-186. (2002). Closely-spaced parallel approaches: Application description version 1.2. Retrieved 26 April 2004, from http://human-factors.arc.nasa.gov/ihi/personnel/ashford/CSPA20D.pdf.

Safety Regulation Group (2002). Human factors in aircraft maintenance and inspection. West Sussex, UK: Civil Aviation Authority.

Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology*, 1(1), 45-57.

Sarter, N. B., & Woods, D. D. (1995). How in the world did I ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37, 5-19.

Selcon, S. J., & Taylor, R. M. (1989). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations*, 5:1-5:19.

Shamo, M., Dror, R., & Degani, A. (1998). Evaluation of a new cockpit device: The integrated electronic information system. *Proceedings of the 42nd Human Factors and Ergonomics Society Meeting*. Chicago, IL: The Human Factors and Ergonomics Society.

Sheridan, T. B. (2002). Some musings on four ways humans couple: Implications for systems design. *IEEE Transactions on System, Man, and Cybernetics – Part A: Systems and Humans*, *32* (1).

Smith, K., & Hancock, P.A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, 37(1), 137-148.

Stone, R. (1996). Paired approach concept. In Waller, M.C. and C.H. Scanlon (Eds.) Proceedings of the NASA Workshop on Flight Deck Centered Parallel Runway Approaches in Instrument Meteorological Conditions. NASA Langley Research Center.

Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations*, 3:1-3:17.

Teo, R. and Tomlin, C. J. (2003). Computing danger zones for provably safe closely spaced parallel approaches. *AIAA Journal of Guidance, Control, and Dynamics*, 26 (3), 434-443.

Trager, E. A. (1988). Special study report: Significant events involving procedures. *AOED/S801*. Washington, DC: Nuclear Regulatory Commission.

Vicente, K. J. & Rasmussen, J. (1992). Ecological interface design: theoretical foundations. IEEE Transactions on Systems, Man, and Cybernetics, SMC-22, 589-6.

Vidulich, M. A. (1992). Measuring situation awareness. In *Proceedings of the Human Factors Society 36th Annual Meeting*, 40-41.

Vita

Steven James Landry

Steven James Landry, the son of Lawrence Joseph Landry and Ruth Virginia (Morse) Landry was born on June 22, 1965 in Milford, MA, and grew up in the neighboring town of Medway, MA, where he attended high school at the Medway Jr.-Sr. High School. He graduated (with Honors) with a B.S. in Electrical Engineering from Worcester Polytechnic Institute in Worcester, MA in 1987, was commissioned as a 2nd Lieutenant in the United States Air Force (USAF), and went on active duty on April 1, 1988 (irony of the date notwithstanding). After pilot training at Williams AFB in Arizona, he was assigned to fly the C-141B with the 18th Military Airlift Squadron, 438th Military Airlift Wing, 21st Air Force, at McGuire AFB, NJ. He flew over 2,000 hours and was upgraded to Aircraft Commander, Instructor, Flight Examiner, and finally the Chief of Squadron Standardization until the closing of the 18th Airlift Squadron, at which time he joined the 30th Airlift Squadron.

After separating from the Air Force in June 1996, he attended M.I.T., working as a research assistant for Dr. Thomas Sheridan under a NASA grant researching air traffic control issues. He graduated from M. I. T. with an S.M. in Aeronautics and Astronautics in the spring of 1999. He then attended the Georgia Institute of Technology in the School of Industrial and Systems Engineering, concentrating in human-machine systems. He received his Ph.D. with a minor and certificate in cognitive science in 2004.